

https://dspace.org/

Molecular characterization of the hepatitis b virus in South Africa

Bowyer, Sheila. M.

2001
https://hdl.handle.net/10539/26649

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

MOLECULAR CHARACTERIZATION OF THE HEPATITIS B VIRUS IN SOUTH AFRICA

SHEILA M. BOWYER

Thesis submitted in compliance with the requirements for the degree of Doctor of Philosophy

in the Department of Health Science at the University of the Witwatersrand

Johannesburg 2001

ACKNOWLEDGEMENTS:

This work was supported by a research grant from the Poliomyelitis Research Foundation.

I would like to thank the following people:

1. Dr. C. Tiemessen, my supervisor, for her support, patience, help, advice and time throughout the project.
2. Dr. J.G.M. Sim for his help and advice in the early years of the project.
3. Dr. L. Morris for reading, advising, editing and helping with the preparation of the manuscript for the recombination study.
4. Dr. E. Vardas for organising the collection of specimens for the study.
5. Drs. E. Song, J. Botha and K. Seidel for providing whole blood and/or serum from patients with acute and/or chronic hepatitis B.

Parts of this work have been published in the following journals:

1. Bowyer, S.M., van Staden, L., Kew, M.C. and Sim, J.G.M. (1997). A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. *J. Gen. Virol.*; **78:** 1719-1729.
2. Bowyer, S.M. and Sim, J.G.M. (2000). Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates. *J. Gen. Virol.*; **81:** 379-392.

Parts of this work have also been presented at the following scientific meetings:

1. Bowyer, S.M. Strain analysis of hepatitis B virus in southern Africa on the basis of restriction endonuclease analysis and sequencing of polymerase chain reaction products. Vth International Congress on the Impact of Viral Diseases in the Developing World,

Johannesburg, South Africa, July, 1995.

2. Bowyer, S.M., van Staden, L., Kew, M.C. and Sim, J.G.M. Molecular characterization of HBV in sera of southern African patients. Molecular Biology of Hepatitis B Viruses, Cold Spring Harbor Laboratories, Cold Spring Harbor, New York, USA, September, 1996.
3. Bowyer, S.M. and Sim, J.G.M. Molecular characterization of hepatitis B virus in sera of southern African patients. Molecular epidemiology and evolutionary genetics annual workshop, Montpellier, southern France, May 1997.
4. Bowyer, S.M. and Sim, J.G.M. Recombination plays a part in the evolution, diversity and pathology of hepatitis B virus. Molecular Biology of Hepatitis B Viruses, University of California, San Diego, California, USA, September, 1998.
5. Bowyer, S.M., Vardas, E. and Sim, J.G.M. Recombination plays a part in the evolution, diversity and pathology of hepatitis B virus. Virology congress of southern Africa. UCT graduate school of Business, Cape Town, South Africa, October 1998.
6. Bowyer, S.M. and Tiemessen, C.T. Regions of the core protein which elicit a class II immune response in acute and chronic HBV patients from southern Africa and their relation to predicted and published linear epitopes of HBV. Molecular Biology of Hepatitis B Viruses, University of Massachusetts, Amherst, Massachusetts, USA, July 2001.

ABSTRACT

The hepatitis B virus (HBV) partitions into 6 genotypes (A-F) and a seventh putative genotype, genotype G, which has recently been described. As southern Africa is highly endemic for HBV infection and little is known of the molecular epidemiology of HBV in this part of the world, we determined the genotype of viral strains circulating in both rural and urban South Africa and rural Namibia. Two screening assays were used. The first was based on size and restriction fragment length polymorphism (RFLP) analysis of specific polymerase chain reaction (PCR) fragments amplified from the preS1 and preS2 regions of the surface gene. The second method was an enzyme immunoassay (EIA) based serological determination of genotype which used monoclonal antibodies raised against genotype-specific epitopes in the preS2-region. Only genotypes A and D were found in significant numbers in South Africa. The urban specimens partitioned as follows: 69.6 percent genotype A, 3.6 percent genotype B, 7.1 percent genotype C, 16.1 percent genotype D and 3.6 percent genotype E. Rural specimens were all either genotype A (89.2 percent) or genotype D (10.8 percent).

Sequence studies showed that many of the African genotype A specimens differed from the well conserved genotype A specimens found in other parts of the world. Phylogenetic studies of sequence data from the preS2 and surface region of the genome showed that these specimens clustered together and away from global genotype A specimens to form a subgroup which we have called subgroup A'. The genotype A specimens in the urban cohort consisted of 35.9 percent subgroup A and 64.1 percent subgroup A' while the rural cohort had only one (3.0 percent) specimen from the A subgroup while the rest (97.0 percent) were from subgroup A'. Two subgroup A' specimens from the Philippines also clustered with this new subgroup and recent data from South America indicates that the subgroup A' is also present in the New World.

Rural Namibia had 93.8 percent genotype E and 6.2 percent genotype D. However, all of the specimens came from the North-West and North-East of the country and these results may not be representative of the whole of Namibia. Pair-wise analysis of local Genotype E sequence

data showed that they differed by less than 4 percent from the two available genotype E specimens from GenBank.

Having established the genotypes into which local specimens partitioned we examined the four open reading frames of HBV at both the nucleotide and amino acid level in order to identify the definitive variation which partitioned the specimens into genotypes and subgroups. At the same time we examined and listed the non-definitive variation between individual specimens within a genotype or genotypic subgroup. Mapping of variation within the virus revealed that the virus is geographically unique in as much as the genotypes and sub-groups partition globally, but there was no significant variation between specimens of the same genotype or sub-group from different parts of the world.

The average evolutionary distance between the type, subtype and isolate categories of HBV specimens was also plotted at each of 54 points along the genome and this revealed that some specimens assigned to different genotypes are more similar across several contiguous intervals than specimens belonging to the same genotype. We found that A/D, A/E and B/C recombination appears to be fairly common in HBV infection and recently several groups have corroborated our findings (Yuasa *et al.*, 2000; Hannoun *et al.*, 2000; Owiredu *et al.*, 2001). Genotypes D and E, as well as genotype C and two of the subgroups of genotype B, do not partition independently in the core gene region. The nature and distribution of polymorphic sites in mosaic regions was mapped at both the nucleotide and protein levels and the position of the variant fragments was related to mutational "hotspots" (where "hotspots" were defined as amino acids which show variation from the consensus in more than 5 percent of specimens examined) and linear epitopes of HBV. The distribution of genetic distances over subgenomic regions showed that substitution rates are not constant among the lineages of HBV in the preS regions. Genotype F is the most diverse genotype. Only genotypes A, C and F partition consistently into subgroups.

Having qualified viral variation we examined the genetic make-up of the African host, mapped immunogenic regions of the core gene product within the African host and compared these local epitopes to predicted and published HBV epitopes. Finally we looked carefully at variation within these epitopes and considered the known or expected effect this variation

would have at the protein level. The core region was chosen for this section of the study since changes in the immune recognition sites of the core antigen have been found to illicit (or evade) immune clearance and affect the outcome of infection.

With the exception of two alleles, the 11 human leucocyte antigen (HLA) DR typed patients had DR alleles common within the Bantu of southern Africa as reported in population studies. There was a predominance of DR03 in our small study. No HLA-DR13 alleles were found in the patients which corroborates the strong association between DRB1*13 and protection from chronic hepatitis B disease. Immunogenic regions of the HBV genome did not differ in local patients although several additional epitopes, specific to the African host, were found. Although indications were that the host HLA type does have an effect on disease progression, the contribution of host factors to HBV disease outcome in southern Africa will be clearer when more high resolution HLA data becomes available.

The precore plus core sequence data from 14 southern African patients (10 of whom were from genotype E, two from genotype C, and one each from genotypes D and the subgroup A') revealed only two specimens (14 percent; C66 and E03) had a guanine to adenine mutation at either nucleotide 1896 or 1897 both of which result in a premature precore stop mutation. These two specimens and one other (C68) had a guanine to adenine switch at nucleotide 1899 which results in a glycine to aspartic acid replacement at the last amino acid of the precore region. Three other specimens (E07, A'03 and E06) had premature stop codons at one or other position of the core which would result in a truncated core antigen. We found the guanine to thymine mutation at nucleotide 1862, which is also thought to disrupt HBeAg expression, to be definitive of subgroup A'. Forty percent of changes in the core gene were found at definitive sites, and approximately 60 percent of both the definitive and the non-definitive changes were predicted to alter the properties of the protein.

As more countries adopt the World Health Organisation guidelines on universal HBV vaccination, the focus of HBV research is shifting to anti-viral therapies including therapeutic DNA and peptide vaccines. For maximum effect the latter require a knowledge of the genetics of the local population. There are, however, many sub-Saharan countries from which HBV molecular epidemiological data is still unavailable and there is still a need for both baseline studies and routine surveillance in this part of the world to identify circulating viral

strains and to prevent the emergence of variants, and anti-viral and vaccine escape mutants in particular. Many recent studies associate viral genotype with disease progression and the evolutionary history of HBV is still unclear possibly because other, as yet unidentified, genotypes and subgroups of both human and primate HBV exist.

DECLARATION

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

(Sheila M. Bowyer)

_____ day of _____, 200_.

For David

with heartfelt thanks for always being there for me, for being interested and supportive in everything I undertake and this project in particular

ABBREVIATIONS:

3D	three dimensional
A	adenine
Acc. No.	GenBank accession number
AFP	alpha feto protein
Ala (A)	alanine
ALT	alanine amino-transferease
anti-HBc	antibody to Hepatitis B core antigen
anti-HBe	antibody to Hepatitis B e antigen
anti-HBs	antibody to Hepatitis B surface antigen
AP-1	activating protein 1
Arg (R)	arginine
ASC	asymptomatic carriers
Asn	antisense
Asn (N)	asparagine
Asp (D)	aspartic acid
ATF-2	activating transcription factor 2
bis	N, N'-methylene-bis-acrylamide
bp	base pairs
C	cytosine
CAH	chronic active hepatitis
cAMP	cyclic adenosine 5'-monophosphate
cccDNA	covalently-closed circular (supercoiled) deoxyribonucleic acid
C/EBP	CCAAT enhancer binding protein
CPH	chronic persistent hepatitis
cpm	counts per minute
CRE	cyclic AMP response element
CREB	cyclic AMP response element binding factor
CTL	cytotoxic T lymphocyte
CURS	core upstream regulatory sequence
Cys (C)	cysteine
DAG	diacylglycerol

DHBV	duck hepatitis B virus
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
dNTPs	deoxynucleoside triphospates
DR1	direct repeat region 1(nucleotide 1824-1834)
DR2	direct repeat region 2 (nucleotide 1590-1600)
ϵ	encapsidation signal
EC	extracellular
EDTA	ethylene diamine tetra-acetate
EF-C	enhanced factor C
EIA	enzyme immunoassay
ENHI	enhancer I
ENHII	enhancer II
env	envelope
ER	endoplasmic reticulum
FCS	foetal calf serum
G	guanine
GAPD	glyceraldehyde-3-phosphate-dehydrogenase
Gln (Q)	glutamine
Glu (E)	glutamic acid
Gly (G)	glycine
gp 180	a 180 kilodalton carboxypepsidase D
GRE	glucocorticoid responsive element
GSHV	ground squirrel hepatitis virus
GTB	glycerol tolerant buffer
^{3}H-thymidine	tritiated thymidine
hAV	human annexin V
HBcAg	hepatitis B core antigen
HBeAg	hepatitis B e antigen
HBsAg	hepatitis B surface antigen
HBxAg or HBx	hepatitis B X antigen
HBV	hepatitis B virus
HBVPRE	hepatitis B virus post-transcriptional regulatory element
HC	hepatocyte
HCC	hepatocellular carcinoma
HHBV	heron hepatitis B virus

His	histidine
HIV	human immunodeficiency virus
HLA	human leucocyte antigen
HNF1/2/3/4	hepatocyte nuclear factor 1/2/3/4
Hsp	heat shock protein
IC	intra-cellular
IFN-γ	interferon gamma
IgM	immunoglobulin M
IHWC	international histocompatibility workshop
IκB	NFκB inhibitor
IL-1/6/8	interleukin-1/6/8
Ile (I)	isoleucine
kb	kilobase
kd	kilodalton
l	litre
LER	liver enriched factor
Leu (L)	leucine
LHBs	large surface protein
LSR	liver specific regulatory element
LTR	long terminal repeat
Lys (K)	lysine
M	Molar
MAbs	monoclonal antibodies
MAPK	microtubule associated protein kinase
MAPKK	microtubule associated protein kinase kinase
Met (M)	methionine
$MgCl_2$	magnesium chloride
MHBs	middle surface protein
MHC	major histocompatibility complex
ml	millilitre
mM	millimolar
mRNA	messenger RNA
NBTS	Namibian blood transfusion service
net	negative effector of transcription
NFκB	nuclear factor κB
ng	nanogram

NIH	National Institutes of Health
NIV	National Institute for Virology
nm	nanometre
NRE	negative regulatory element
nt	nucleotide
Oct-1/2	immunoglobulin gene octamer motif binding protein
ORF	open reading frame
P	the polymerase ORF also called *Pol*
PBA	polyclonal B cell activator
PBMCs	peripheral blood mononuclear cells
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pet	positive effector of transciption
pgRNA	pregenome ribonucleic acid
PHA	phytohaemagglutinin
Phe (F)	phenylalanine
PKA	protein kinase A
PKC	protein kinase C
pol	polymerase
polyA	polyadenylation signal
PPAR	peronisome proliferator-activated receptor
PRE	post-transcriptional regulatory element
Pro (P)	proline
RARE	retinoid acid responsive element
rbcs	red blood corpuscles
RC DNA	relaxed circular deoxyribonucleic acid
RE	restriction endonucleases
RF1/2/3	reading frame 1/2/3
RFLP	restriction fragment length polymorphism
RFX	regulatory factor X
RIA	radioimmunoassay
RNA	ribonucleic acid
RNAse H	ribonuclease H
RNP	ribonucleoprotein
ROI	reactive oxygen intermediates
RT	reverse transcriptase
RXRα	retinoid X receptor alpha

SD	standard deviation
SDS	sodium dodecyl sulphate
SE	standard error
Ser (S)	serine
SHBs	small surface protein also referred to as HBsAg
SI	stimulation index
Sn	Sense
SPI	preS1 promoter
SPII	preS2/S promoter
SRE	sterol responsive elemen
SSCP	single-stranded conformational polymorphism
SSP	sequence specific primer
T	thymine
TBB	Tris-Borate buffer
TBP	TATA-binding protein
TCR	T-cell receptor
Th	T helper cell
Thr (T)	threonine
TNFα	tumour necrosis factor alpha
Top I	topoisomerase I
TP	terminal protein
TPA	12-o-tetradecanol phorbol-13 acetate
Tris	Tris(hydroxymethyl) aminomethane
Tris-HCl	Tris hydrochloride
Try (W)	tryptophan
Tyr (Y)	tyrosine
UE1/2	unidentified protein binding site
μg	microgram
μl	microlitre
μM	micromolar
URR	upstream regulatory region
UV-DDB	ultra violet-damaged DNA binding protein
V	Volts
Val (V)	Valine
WHO	World Health Organisation
WHV	Woodchuck hepatitis virus
WMHBV	Woolly monkey hepatitis B virus

LIST OF TABLES

Table 1.	Structural and functional properties of the HBV transcripts	4
Table 2.	Published class I- and class II-restricted epitopes of HBV	25
Table 3.	The three serological profiles of HBV chronic carriage	29
Table 4.	(a) HBV specific primer sequences and (b) resultant amplicon sizes obtained using appropriate forward and reverse primer pairs	43
Table 5.	Sn and Asn sequencing primers used for sequencing the surface and core genes	46
Table 6.	The position and nature of the overlapping core gene peptides	48
Table 7.	Prevalence of HBV serological markers in sub-Saharan Africa	54
Table 8.	Different published methods of HBV serotyping and/or genotyping	61
Table 9.	Size and restriction fragment length polymorphisms used to partition isolates into the 6 major genotypes and two subtypes	63
Table 10.	Summary of results from South African specimens genotyped by RFLP analysis and size polymorphism, HBV genotype-specific EIA or sequence analysis	68
Table 11.	Summary of results from Namibian specimens genotyped by RFLP analysis and size polymorphism, HBV genotype-specific EIA or sequence analysis	71
Table 12.	Summary of genotype screening results from Namibia	70
Table 13.	Designation of 110 HBV isolates used to characterize the subgroup A' of genotype A	78
Table 14.	Percentage diversity at the nucleotide and amino acid level across the different domains spanning the preS2/S gene open reading frame	80
Table 15.	Similarity of fragments derived from mosaic areas to parental subgroups	130

LIST OF TABLES (continued)

Table 16.	A short description of some African populations studied and codes of ethnic groups with which they were compared	147
Table 17.	(a) The major splits of the HLA-A, -B and -DR antigens. (b) The five haplotypes and 13 lineages of HLA-DR. (c) The HLA-A and -B alleles grouped into the 9 recognised supertypes	148
Table 18.	HLA-DR types determined for southern African specimens	157
Table 19.	Published host susceptibility genes which modify the disease phenotype, the vaccine response or the response to therapy	159
Table 20.	Population frequencies of the HLA-DR alleles in four Bantu populations and the San people	160
Table 21.	Genotype and HLA-DR type of specimens chosen for proliferation analysis	166
Table 22.	HLA-binding motifs present in predicted and published epitopes	169
Table 23.	Proliferation assay responses of PBMCs to overlapping synthetic HBV core peptide analogues	170
Table 24.	Amino acid changes (and their possible effects) in the HBc/e protein sequence	195-197
Table 25.	GenBank and local (sequenced) patient details	219-229
Table 26.	Definitive and quasi-species amino acid differences over the ORFs of HBV	253-265
Table 27.	(a) Preliminary counts used to determine optimum concentration of PHA and the effect of peptide solvents on cell proliferation (b) Mean counts of ^{3}H-thymidine incorporated into PBMCs in response to stimulation by mitogen or overlapping core antigen peptides in acute and chronic hepatitis	266-272
Table 28.	Summary of source, size and significance of samples used in (a) Determination of molecular epidemiology of the different genotypes of HBV in rural and urban South Africa and the rural areas of Northern Namibia (b) Molecular epidemiology of a subset of 29 South African specimens genotyped by sequence analysis	275

LIST OF FIGURES

Figure 1.	Map of the hepatitis B virus genome showing *cis* regulatory elements, translated regions and mRNA start sites	3
Figure 2.	Model of the core antigen (HBcAg) showing core protein domains	9
Figure 3.	Model of HBsAg in the membrane showing the "a" epitope and the glycosylation site	11
Figure 4.	The two different transmembrane topologies adopted by the large surface protein (a) e-preS in which the preS1 domain is exposed on the surface and (b) i-preS with internal preS1 domain	13
Figure 5.	The functional domains of *pol* shown in the context of the full-length HBV polymerase	14
Figure 6.	Schematic representation of the characteristic regions in the HBV X protein	16
Figure 7.	Some of the pathways which may be altered by HBx	18
Figure 8.	The life cycle of HBV	21
Figure 9.	Replication strategy of HBV	23
Figure 10.	Serological and clinical patterns associated with HBV infection (a) with resolution and (b) without resolution	27
Figure 11.	Model of HBV immunopathogenesis	33
Figure 12.	The HBV PCR reaction with primers Sn3088 and Asn256 results in amplification of a 413 bp fragment which includes the entire preS2 region	44
Figure 13.	HLA-DR specific bands for patient ES06	50
Figure 14.	High resolution HLA-DR3 specific bands for patient HB05	51
Figure 15.	Prevalence of hepatitis B virus infection worldwide	53
Figure 16.	Percentage prevalence of HBsAg in various towns and regions of southern Africa	55
Figure 17.	World distribution of HBsAg serotypes	57

LIST OF FIGURES (continued)

Figure 18.	World distribution of HBV genotypes/serotypes	59
Figure 19.	Restriction endonuclease sites used in the RFLP genotype screening assay	63
Figure 20.	Serological partitioning of isolates into genotypes A, B, C and F as well as a fifth partition made up of specimens from both genotype D and E using HBV genotype-specific EIA	65
Figure 21.	PS1 amplicons on 3 percent Metaphor agarose showing the clear distinction between the 128 bp amplicon of genotype D and the larger (158-161 bp) amplicons of all other genotypes	67
Figure 22.	PS1 *Bam HI* digestion	67
Figure 23.	Phylogenetic distance tree produced using PHYLIP programmes FITCH and DRAWTREE from preS2/S sequence data from 26 South African sequences and 42 global sequences from GenBank	81
Figure 24.	Phylogenetic trees produced by the PHYLIP non-distance matrix programme DNAML using preS2/S sequence data. (a) 21 South African and 10 GenBank genotype A sequences and (b) three South African and 10 GenBank group D sequences	82
Figure 25.	Mean intra- and intergroup percentage divergence in nucleotide and amino acid sequences of the preS2/S gene region of (a) the 6 genotypic groups A to F and (b) the two subgroups of A	83
Figure 26.	PreS2/S amino acids present at HBV variant sites in representative specimens from genotypes A to F	85-86
Figure 27.	Flowchart of the computer programmes written to analyse the 65 full-genome sequences from GenBank together with the filenames of the databases they use as input or generate as output	93
Figure 28.	Frequency of the consensus nucleotide of 65 full-genome HBV sequences from GenBank plotted at each position in the genome	94
Figure 29.	Frequency of the consensus amino acid of 65 full-genome HBV sequences from GenBank plotted at each position in the HBV genome in (a) RF1 (b) RF2 and (c) RF3	96

LIST OF FIGURES (continued)

Figure 30.	The "hotspots" in ORF1 of RF1 which codes for the preS2 and HBsAg domains	97
Figure 31.	The "hotspots" in ORF2 of RF1 coding for the precore and core domains	98
Figure 32.	The "hotspots" on the X protein in RF2	100
Figure 33.	The "hotspots" on the TP domain of the P gene in RF2 including the spacer region in this reading frame	101
Figure 34.	The "hotspots" on the RT/*Pol* and RNAse H domains of the P gene in RF3	103
Figure 35.	The "hotspots" on the preS1 domain of the S gene in RF3	104
Figure 36.	The number of definitive versus quasi-species amino acid changes in RF1 coded regions of 65 specimens grouped by genotype in (a) the preS2 domain (b) the HBsAg region and (c) the precore/core region	106-108
Figure 37.	The number of definitive versus quasi-species amino acid changes in RF2 coded regions of 65 specimens grouped by genotype in (a) the X region and (b) the TP domain of P	109-110
Figure 38.	The number of definitive versus quasi-species amino acid changes in RF3 coded regions of 65 specimens grouped by genotype in (a) the RT and (b) the RNAse H domain of *Pol* and (c) the preS1 domain of *env*	110-112
Figure 39.	Quasi-species change per amino acid in each of 65 full-genome specimens in the precore/core domain	115
Figure 40.	Full-genome FITCH bootstrapped consensus tree (100 sets) showing the conventional segregation of HBV specimens into the 6 genotypic groups (A-F)	120
Figure 41.	Interval number (on the x axis) is plotted against average genetic distance: (a) within subgroups, (b) between subgroups of the same genotype, and (c) between selected subgroups of different genotypes	124-127
Figure 42.	Graphs of pair-wise distance of putative recombinants from parental subgroups to illustrate breakpoint determination. Average pair-wise distances over each of the 54 intervals between (a) isolate Dm16 and i. subgroup A and ii. subgroup D and (b) isolate B'01 and i. subgroup C and ii. subgroup B	129

LIST OF FIGURES (continued)

Figure 43	The distribution and nature of variation within individual isolates mapped against the consensus sequence of the parental subtypes as described in the text. (a) Mosaic region (nucleotide 822-1775) of specimen Dm17. For clarity, only variant sites are listed. (b) By coding each batch of 20 nucleotides in the same way as before, the complete genome, and not just variant positions, of 13/14 putative recombinants and one non-recombinant, A'02, is mapped proportionately to show mosaic areas in context	131-133
Figure 44.	Bootstrap resampling trees of 4/9 mosaic fragments. (a) Fragment II, intervals 9-13, present in Dm16 only. (b) Fragment IV, intervals 15-30 is present in Dm17 only but includes parts of fragments III and V of Dm18. (c) Fragment VI, intervals 27-39, present in E01 and E02. (d) Fragment VIII, intervals 30-37, present in B'01 to B'04 and both B"11 and B"12 also includes part of fragment V of Dm18	134
Figure 45.	The distribution of pair-wise genetic distances (a) over each of 7 named subgenomic genes/domains, the intervals covered by each region are shown in parentheses, or (b) over the complete genome	136
Figure 46.	The distribution of languages from the basic African Linguistic groups in Africa today	142
Figure 47.	Distribution of major ethnic groups in southern Africa today	145
Figure 48.	Percentage frequency of major HLA-A antigens in selected populations	149
Figure 49.	Percentage frequency of major HLA-B antigens in selected populations	152
Figure 50.	Percentage frequency of major HLA-DR antigens in selected populations	154
Figure 51.	Percentage frequency of major HLA-DQ antigens in selected populations	156
Figure 52.	Published epitopes in the HBV core protein superimposed on the histogram of MHC-binding motif density as predicted by BIMAS and SYFPEITHI	167
Figure 53.	Stimulation index plotted against peptide used to prime PBMCs from representative patients with (a) acute hepatitis B (i. and ii.) and (b) chronic hepatitis B (i. and ii).	171

LIST OF FIGURES (continued)

Figure 54.	Published class I and class II epitopes and their position within motif-rich regions	172
Figure 55.	Distribution of A2 subtypes in selected sub-Saharan Africans and Europeans	178
Figure 56.	Neighbour-joining tree (UPGMA option) of 175 core gene sequences	182
Figure 57.	Average nucleotide differences between specimens from genotypes A to G and an additional group, Ch, which contains four specimens from the newly identified chimpanzee hepatitis B virus.	183
Figure 58.	Average genetic distance between the genotypes specified plotted against the first nucleotide number of the interval (with 1 representing the start of the precore) over the 645 nucleotides of the precore/core gene. In all graphs the core gene can be seen to be made up of two conserved regions sandwiched between a more variable region. (a) Genotypes D and E have only intra-genotypic differences over most of the core gene, (b) subgroup B' of genotype B is more closely related to the two subgroups of genotype C and subgroup B" than it is to the third subgroup of genotype B, subgroup B, and (c) only normal inter-genotypic differences are observed between genotype E and genotype F and genotype E and the new genotypes, genotype G and the chimpanzee hepatitis B virus	184
Figure 59.	Variation within published epitope core-1-20. The twelve amino acid insert definitive of genotype G is within this epitope	185
Figure 60.	Variation within the published class I HLA A2-restricted epitope core-18-27	186
Figure 61.	Variation within the published epitope core-28-47	187
Figure 62.	Variation within the published class II -restricted epitope core-50-69	188
Figure 63.	Variation within the published class I HLA A11-restricted epitope core-87-100	189
Figure 64.	Variation within the published class II DR1-restricted epitope core-111-125	190
Figure 65.	Variation within the published class II DR52/DR6-restricted epitope core-117-131	190

LIST OF FIGURES (continued)

Figure 66.	Variation within the published class II DR1-restricted epitope core-120-139	190
Figure 67.	The quasi-species and genotypic change within 175 core genes showing defined B and T-cell regions, HLA class I and II motif-dense regions and the 7 heterogenous domains (I-VII) defined by Hosono *et al.* (1995)	191
Figure 68.	Variation within the published class II DR1-restricted epitope core-141-151	192
Figure 69.	Evolution of modern primates	208

TABLE OF CONTENTS

Title	i
Acknowledgements	ii
Abstract	iv
Declaration	viii
Dedication	ix
Abbreviations	x
List of Tables	xv
List of Figures	xvii
Table of contents	xxiii
Note re numbering of the hepatitis B virus genome	xxxiv
Chapter 1 Literature Review	1
1.1 The hepatitis B virus	1
1.1.1 The viral genome	1
1.1.2 Transcription	2
1.1.2.1 Class I: The 3.5 kb transcripts	3
1.1.2.2 Class II: The 2.1 kb transcripts	5
1.1.2.3 Class III: The 2.4 kb transcript	5
1.1.2.4 Class IV: The 0.7 kb transcript	5
1.1.2.5 Spliced transcripts	6
1.1.2.6 Regulation of HBV RNA export	6
1.1.2.7 3' end formation	7
1.1.3 The structure and function of the 7 HBV proteins	7
1.1.3.1 The core (*gag*) gene products	8

TABLE OF CONTENTS (continued)

1.1.3.1.1 The core antigen, HBcAg 8

1.1.3.1.2 The e antigen, HBeAg 9

1.1.3.2 The surface (*env*) gene products 10

1.1.3.2.1 The small (major) surface antigen, SHBs (HBsAg) 11

1.1.3.2.2 The middle surface protein, MHBs 12

1.1.3.2.3 The large surface protein, LHBs 12

1.1.3.3 The P (*pol*) gene 14

1.1.3.4 The X gene product, HBx 15

1.2. The immunopathogenesis of HBV 19

1.2.1 The life cycle of HBV 19

1.2.1.1 Viral binding and cell entry 19

1.2.1.2 cccDNA synthesis and integration 20

1.2.1.3 Viral replication 22

1.2.1.4 Packaging and export 24

1.2.2 Outcomes of HBV infection 24

1.2.2.1 Acute self limiting hepatitis B 24

1.2.2.2 Chronic HBV infection 26

1.2.2.3 Hepatocellular carcinoma, HCC 29

1.2.3 Common mutants of HBV and their effect on pathogenesis 31

1.2.3.1 Precore mutants 31

1.2.3.2 Core promoter mutants 31

1.2.3.3 PreS1 deletion mutants 32

TABLE OF CONTENTS (continued)

1.2.3.4 Core internal deletion mutants 32

1.2.3.5 Variants containing spliced genomes 33

1.2.3.6 Surface antigen escape mutants 33

1.2.4 Model of HBV infection 33

Chapter 2 Introduction: Rationale for the Study 36

2.1 Motivation for the study 36

2.2 Aims and organisation of the study 38

Chapter 3 Materials and Methods 40

3.1 Patients 40

3.1.1 South African patients 40

3.1.1.1 Urban cohort 40

3.1.1.2 Rural cohort 40

3.1.2 Namibian patients 41

3.1.3 GenBank specimens 41

3.2 Molecular methods 42

3.2.1 Viral DNA extraction 42

3.2.2 Polymerase chain reaction (PCR) 42

3.2.3 Restriction endonuclease (RE) reactions 45

3.2.4 Sequencing of PCR products 45

3.3 Immunological methods 47

3.3.1 Lymphocyte preparation 47

3.3.2 Antigen preparation and optimization 48

3.3.3 Cell culture conditions 49

3.3.4 Proliferation assay 49

TABLE OF CONTENTS (continued)

3.3.5 Class II (HLA-DR) typing 50

Chapter 4 The Epidemiology of HBV Infection 52

4.1 Introduction 52

4.1.1 Epidemiology of HBV infection in southern Africa 53

4.1.1.1 Regional prevalence studies in southern Africa 53

4.1.2 Serotypes of HBV 56

4.1.3 Genotypes of HBV 58

4.1.4 Molecular epidemiological studies and strategies 60

4.2 Materials and Methods 60

4.2.1 Patients 60

4.2.2 Genotype screening 62

4.2.2.1 Rationale for RFLP and size polymorphism analysis 62

4.2.2.2 PCR amplification of HBV DNA 64

4.2.2.3 Rationale for HBV genotype-specific EIA 64

4.3 Results 66

4.3.1 Size polymorphism results 66

4.3.2 South African genotyping results 66

4.3.3 Namibian genotyping results 70

4.4 Discussion 72

Chapter 5 A unique segment of HBV genotype A identified in isolates from South Africa 75

5.1 Introduction 75

5.2 Materials and Methods 75

TABLE OF CONTENTS (continued)

5.2.1 Patients 75

5.2.2 HBV DNA extraction 76

5.2.3 DNA amplification 76

5.2.4 DNA sequencing 76

5.2.5 Serotyping 77

5.2.6 Data analysis 77

5.3 Results 77

5.4 Discussion 84

Chapter 6 Mapping the "hotspots", definitive sites and distribution of individual variation on the HBV genome 90

6.1 Introduction 90

6.2 Materials and Methods 91

6.2.1 Patients 91

6.2.2 Database analysis 92

6.2.2.1 Nucleotide and amino acid variation 92

6.2.2.2 Rapid alignment 92

6.3 Results 92

6.3.1 The frequency and position of "hotspots" 92

6.3.1.1 "Hotspots" within the coding regions of RF1 95

6.3.1.2 "Hotspots" within the coding regions of RF2 99

6.3.1.3 "Hotspots" within the coding regions of RF3 102

TABLE OF CONTENTS (continued)

6.3.2 Individual change within specific isolates 105

6.3.2.1 Non-definitive change in RF1 106

6.3.2.2 Non-definitive change in RF2 108

6.3.2.3 Non-definitive change in RF3 109

6.4 Discussion 111

Chapter 7 Relationships within and between genotypes of HBV at points across the genome: Footprints of recombination in certain isolates 118

7.1 Introduction 118

7.2 Materials and Methods 119

7.2.1 Patients 119

7.2.2 Phylogenetic analyses 119

7.2.3 Average genetic distance graphs 119

7.2.4 Isolate screening 121

7.2.5 Individual genetic distance graphs 121

7.2.6 Nucleotide/protein maps 121

7.2.7 Subgenomic bootstrap trees 122

7.2.8 Histograms 122

7.3 Results 122

7.3.1 Subgroups of HBV 122

7.3.2 Average genetic distance graphs 123

TABLE OF CONTENTS (continued)

7.3.2.1 Intra-subgroup 123

7.3.2.2 Inter-subgroup 123

7.3.2.3 Inter-genotype 125

7.3.3 Mosaic structure within isolates 126

7.3.4 Effect of parallel evolution 128

7.3.5 Effect of host selective pressure 131

7.3.6 Effect of functional constraint 131

7.3.7 Histograms 131

7.4 Discussion 135

Chapter 8 HLA frequency and the susceptibility to or protection from HBV in southern Africa 141

8.1 Introduction 141

8.2 Materials and Methods 145

8.2.1 Patients 145

8.2.2 Populations 145

8.2.3 Low resolution HLA-DR typing of southern African patients 146

8.3 Results 146

8.3.1 Class I HLA-A comparison 146

8.3.2 Class I HLA-B comparison 151

8.3.3 Class II HLA-DR comparison 153

8.3.4 Class II HLA-DQ comparison 155

8.3.5 Local HLA-DR typing results 157

8.4 Discussion 158

TABLE OF CONTENTS (continued)

Chapter 9 Mapping the regions of the HBV core gene product which elicit an immune response in acute and chronic patients from southern Africa 163

9.1 Introduction 163

9.2 Materials and Methods 164

9.2.1 Study populations 164

9.2.2 Serology 164

9.2.3 Genotype determination 165

9.2.4 Low resolution HLA-DR typing of southern African patients 165

9.2.5 Lymphocyte isolation and proliferation assay 165

9.2.6 Peptide preparation 165

9.2.7 Expression of data 165

9.2.8 Epitope prediction 166

9.3 Results 166

9.3.1 Genotypes 166

9.3.2 HLA-DR typing 167

9.3.3 Regions predicted to contain the HBV linear epitopes 167

9.3.3.1 The motif-rich regions in relation to published class I epitopes 168

9.3.3.2 The motif-rich regions in relation to published class II epitopes 168

9.3.4 Proliferation responses to the core peptide analogues 168

9.4 Discussion 175

Chapter 10 HBV core gene variation within predicted and published epitopes 180

10.1 Introduction 180

10.2 Materials and Methods 180

10.2.1 Patients 180

TABLE OF CONTENTS (continued)

10.2.2 Sequencing 181

10.2.3 Phylogenetic analysis 181

10.3 Results 181

10.3.1 Phylogenetic analysis 181

10.3.2 Core type-relationship graphs 183

10.3.3 Variation within the predicted and published epitopes 185

10.3.3.1 The genotype G insert 185

10.3.3.2 Potential T-cell epitope region I (amino acid 635-672) 185

10.3.3.3 Potential T-cell epitope region II (amino acid 684-705) 187

10.3.3.4 Potential T-cell epitope region III (amino acid 709-761) 188

10.3.3.5 Potential T-cell epitope region IV (amino acid 765-764) 192

10.3.3.6 Potential T-cell epitope region V (amino acid 778-785) 192

10.3.3.7 Potential T-cell epitope region VI (amino acid 798-802) 192

10.4 Discussion 193

Chapter 11 Final discussion and future direction 199

Appendix A The position of ORFs and binding sites on the HBV genome shown on GenBank sequence accession number X02763 209

Appendix B GenBank and local (sequenced) patient detail 219

Appendix C Computer programmes to determine the "hotspots" on the HBV genome 230

C.1 Reading data into the database 230

TABLE OF CONTENTS (continued)

C.1.1 INPUT.PRG 230

C.1.2 Structure of the nucleotide sequence database BASICDB.DBF 231

C.2 Translating a database of sequence data and storage of the predicted amino acid sequences 231

C.2.1 RF1.PRG 231

C.2.2 RF2.PRG 233

C.2.3 RF3.PRG 235

C.2.4 Structure of the three protein databases PROT1/2/3.DBF 238

C.2.5 FILLAA.PRG 238

C.2.6 Structure of the reformatted protein databases AALIST1/2/3.DBF 240

C.3 Determining the consensus sequence of the records in a database 241

C.3.1 AGCT.PRG 241

C.3.2 Structure of database containing base counts SEQCT.DBF 242

C.3.3 AACT.PRG 243

C.3.4 Structure of database containing amino acid counts RF1/2/3CT.DBF 246

C.3.5 GETMAXNT.PRG 240

C.3.6 Structure of database containing the nucleotide consensus SEQMAX.DBF 241

C.3.7 GETMAXAA.PRG 241

C.3.8 Structure of database containing amino acid consensus AA1/2/3.DBF 242

Appendix D Definitive and quasi-species amino acid differences over the ORFs of HBV 243

TABLE OF CONTENTS (continued)

Appendix E Mean peak counts of ^{3}H thymidine incorporated into peripheral blood mononuclear cells in response to mitogen and HBV core antigen in acute and chronic hepatitis 266

Appendix F Summary of source, size and significance of samples used in each of the seven sections of the study 273

F.1 Molecular epidemiology of HBV infection 273

F.1.1 Calculation of suitable sample size 273

F.1.2 Ninety-five percent confidence interval for each population proportion 274

F.2 Sequence characterization of subgroup A' 274

F.3 Hotspots on the HBV genome and the relationship between the genotypic groups of HBV 274

F.4 HLA frequency and the susceptibility to or protection from HBV in southern Africa and mapping of the immunogenic regions of the core gene 276

F.5 HBV core gene variation 276

References 277

Note re numbering of the hepatitis B virus genome

Since the hepatitis B virus (HBV) has a circular genome which is read three times, once in each of the three possible reading frames, it is important to be clear at the outset which numbering system is being used to refer to the nucleotides and amino acids of the genome. By convention, many people use the unique *Eco RI* site (G↓AATTC), present in most HBV genomes, as the start of numbering. However, this can still lead to confusion since the code is read continuously and the start of coding shifts with each of the three reading frames. In reading frame 1 (RF1) translation is in the frame $G^{1}A^{1}ATTC$, in reading frame 2 (RF2) it is $GA^{321}A^{1}TTC$ and in reading frame 3 (RF3) $GAA^{321}T^{1}TC$. Appendix A provides a visual reference to the molecular features of the virus and their relationship to one another. Numbering in appendix A is in RF3, $GAA^{321}T^{1}TC$. When not otherwise indicated, numbering in the text will be in accordance with the nucleotide numbering in Appendix A. However, when numbering refers to the start of a particular reading frame this is stated and/or indicated clearly.

CHAPTER 1

LITERATURE REVIEW

1.1 The hepatitis B virus

Hepatitis B virus (HBV) is the prototype of the hepadnavirus family, which includes, amongst others, related hepatitis viruses infecting the woodchuck (woodchuck hepatitis virus, WHV), the ground squirrel (ground squirrel hepatitis virus, GSHV), the Pekin duck (duck hepatitis B virus, DHBV), the grey heron (heron hepatitis B virus, HHBV, Summers and Mason, 1982a) and more recently the woolly monkey (woolly monkey hepatitis B virus, WMHBV, Lanford *et al.*, 1998).

1.1.1 The viral genome

The hepadnaviral genome is a partially single-stranded circular deoxyribonucleic acid (DNA) molecule. The extent of the single-stranded portion varies from 15 to 50 percent of the genome length in different molecules (Summers *et al.*, 1975; Landers *et al.*, 1977; Hruska *et al.*, 1977) although the favoured minimum length is 650 to 700 bases ± 20 percent (Delius *et al.*, 1983). The coding of the virus occurs on one strand (Galibert *et al.*, 1979) and the sense direction of the nucleotide sequence was found to be 5' to 3' on the incomplete strand, making it the plus, non-coding strand (Galibert *et al.*, 1982). The HBV genome consists of 3182-3221 base pairs (bp), having the smallest genome of any known animal virus (Robinson *et al.*, 1974). Despite its size the genome is complex and is translated in all three reading frames. Viral regulatory elements are within the HBV protein coding regions which places constraints on the ability of the virus to tolerate mutations and remain viable.

The viral polymerase is covalently linked to the 5' end of the complete DNA minus strand (Gerlich and Robinson, 1980) which prevents phosphorylation of the strand. The complete strand of the viral DNA is thus nicked at a unique site (nucleotide position 1826, Tiollais *et al.*, 1985). At this point the minus strand is redundant by up to 8 nucleotides and forms a short triple-stranded region of the genome (Will *et al.*, 1987).

A substantial number of DNA plus strands carry at their 5' end a ribonucleic acid (RNA) primer about 17 nucleotides in length. This end of the plus strand has been mapped to position 1600 ±1 and is 225 bp from the beginning of the triple-stranded region. The region between the nicks forms an overlap between the two strands which is bound on both sides by an 11 bp direct repeat (DR) sequence (5'-TTCACCTCTGC). One copy of this sequence, termed DR1, starts at nucleotide 1824 while the other, DR2, starts at nucleotide 1590 (Figure 1 and Appendix A). These direct repeats have a biological role in replication and are well conserved in all the hepadnaviruses (Tiollais *et al.*, 1985).

1.1.2 Transcription

Hepadnaviruses multiply their DNA genome by reverse transcription of an RNA intermediate, termed the pregenomic RNA (pgRNA; Summers and Mason, 1982b). As a result hepadnaviral transcription has a dual function: The production of the RNA template for genome replication and the synthesis of messenger RNAs (mRNAs).

Hepadnaviral transcription is highly tissue- specific and no cell culture system exists that is permissive for all steps of the replication cycle. For this reason, information concerning HBV transcription was gleaned initially from physical mapping of the RNA species produced by animal hepadnaviruses in the liver of their respective hosts (Büscher *et al.*, 1985; Enders *et al.*, 1985; Moroy *et al.*, 1985), and on analogous results from studies of an HBV infected chimpanzee (Cattaneo *et al.*, 1983). These experiments, together with site-directed mutagenesis studies, identified numerous hepadnaviral *cis*-acting control elements, pinpointed the exact position of transcription initiation sites in the HBV genome and clarified the structure and coding capacity of individual HBV transcripts (Schaller and Fischer, 1991).

The regulatory regions of cellular genes stretch over a distance of 3 to 10 kilobases (kb) which is longer than the length of the HBV genome. The small circular genome thus behaves as a single complex 'regulon' in which regulatory proteins, binding to any given *cis*-acting element, productively interact, directly or indirectly, with several other such sites or exert effects in combination with other elements (Schaller and Fischer, 1991). As well as four functionally-defined promoters and a single polyadenylation site (polyA), the HBV genome has two liver-specific enhancer regions and a glucocorticoid responsive element

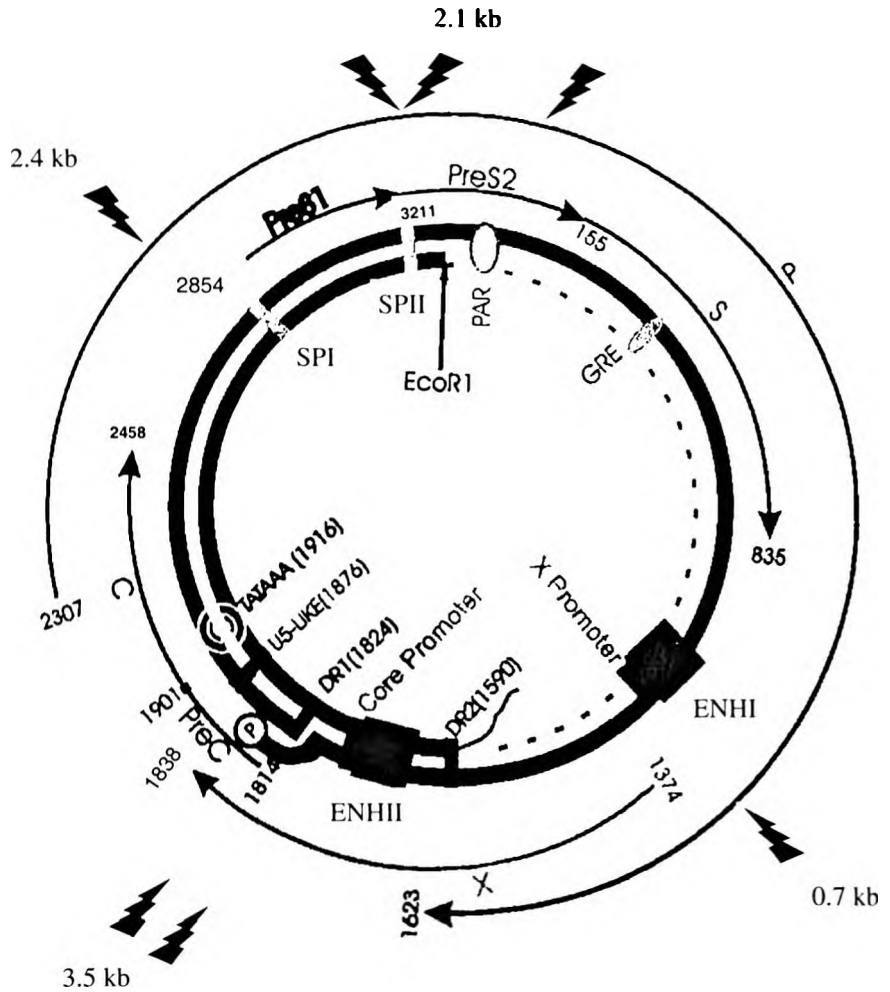


Figure 1. Map of the hepatitis B virus genome showing *cis* regulatory elements, translated regions and mRNA start sites.

Note the four sizes of transcripts: the 3.5 kb transcript under the control of the core promoter and Enhancer II, ENHII, is both the precore and core mRNA as well as the pregenome, the 2.4 kb transcript which is under the control of the preS1 promoter, SP1, and is the mRNA for LHBs, the 2.1 kb transcript which is under the control of the preS2/S promoter, SPII, and is mRNA for both MHBs and HBsAg and lastly the 0.7 kb transcript under the control of the X promoter and Enhancer I, ENHI, which is X protein mRNA. Also shown are the TATA box, the U5-like element, the two direct repeat regions, DR1 and DR2, the glucocorticoid receptor, GRE, and the polymerised albumen receptor domain, PAR.

(GRE). Transcription is controlled by ubiquitous and liver-specific transcription factors, and is modulated in a complex manner by several of the viral proteins. HBV produces four categories of 6 unidirectional and overlapping transcripts from four open reading frames (ORFs) and these are translated into at least 7 viral gene products (Table 1).

1.1.2.1 Class I: The 3.5 kb transcripts

The core promoter (or genomic promoter) is liver-specifically controlled and directs the synthesis of two closely related, but functionally very different, subsets of genomic transcripts which have been designated according to their function as:

1. The precore mRNA and
2. The core mRNA or pregenome.

Table 1. Structural and functional properties of the HBV transcripts

Size (kb)	Start sites	Length 5' to polyA (kb)	Function	Protein	Protein size
3.5	1783/ 1784 or 1790 ± 1	3.35	preCore mRNAs	HBeAg	p22
3.5	1818	3.30	Core mRNAs/ pregenome/ P mRNA	HBcAg RT/*Pol* RNAse H	p21.5 ? ?p35
2.4	2807	2.30	preS1 mRNA	LHBs	p45
2.1	3157/ 3158	2.00	preS2 mRNA	MHBs	p33
2.1	3174/3175 5/6 113	1.95	S mRNA	HBsAg	p24
0.7	1226, 1242	0.65-0.7	X mRNA	HBx	p16.5

These greater than genome length RNAs are terminally redundant and form the most abundant transcripts with a size of about 3.5 kb (Yaginuma *et al.*, 1989). The HBV core promoter is *trans*-activated by human nuclear factor 4 (HNF4), a liver enriched transcription factor, through an upstream HNF4 binding site (nucleotide 1603-1617, Appendix A). This HNF4 site is important for enhancer II (ENHII) activity.

The *trans*-activation activity of HNF4 on the core promoter is antagonized by the negative regulatory element, NRE, which is located upstream of the HNF4 binding site. NRE can be divided into three subregions, NREα (1452-1516), NREβ (1516-1573) and NREγ (1573-1623), each of which bind different protein factors, which explains the differential regulation of HBV according to cell type (Chen and Ou, 1995).

The core mRNA or pregenome is slightly shorter and more abundant than the precore mRNA. It serves as the pregenome as well as mRNA for synthesis of core protein (HBcAg) and, by internal translation initiation, of the polymerase, P or *pol*, protein. The 5' end of the pregenomic RNA is unique and located 5 nucleotides downstream of the precore initiator in HBV at nucleotide 1818 (Appendix A, Will *et al.*, 1987). Assuming a polyA tail of 100 nucleotides, the pregenome is 3.4 kb in length and is terminally redundant by 120 nucleotides. The duplicated sequence elements of the terminal redundancy are essential for

genome circularization and replication. A short 5'-proximal region on the RNA folds to form a characteristic stem-loop structure which serves as an encapsidation signal (ϵ).

The slightly longer precore mRNAs initiate 20 to 30 nucleotides upstream of the precore ORF at either nucleotide 1783/1784 or nucleotide 1789 (Yaginuma *et al.*, 1989), and serve as mRNAs for the synthesis of a fused precore/core gene product, the e antigen (HBeAg), which is secreted and has different structural and antigenic properties to those of HBcAg. Since translation of the precore ORF interferes with genomic RNA packaging, precore mRNAs are not packaged into nucleocapsids. The 5' signal on these transcripts is inactivated by translating ribosomes (Nassal *et al.*, 1990).

1.1.2.2 Class II: The 2.1 kb transcripts

The 2.1 kb subgenomic transcripts are also abundant. They are under the control of the preS2/S promoter, SPII. SPII does not have a TATA transcriptional phasing signal but has GC-rich proximal promoter elements upstream of the CAP site of the preS2 mRNA (Cattaneo *et al.*, 1983). This promoter is not liver-specific and controls the synthesis of several RNA species initiating between positions 3213/3214 and 113 which serve for the expression of the small, or major surface protein, p24 (SHBs or HBsAg). SPII also controls the synthesis of a longer transcript, initiating at positions 3196/3197, which includes code for the preS2 domain and allows translation of the middle surface protein, p33 (MHBs).

1.1.2.3 Class III: The 2.4 kb transcript

The 2.4 kb transcript is a relatively minor RNA and correlates with the synthesis of a third, minor envelope protein termed preS1 or the large envelope protein, LHBs. The preS1 promoter, SPI, is highly liver-specific and the least active of the HBV promoter elements. It has a TATA motif at nucleotide 2782, thirty nucleotides downstream of the CAP site. The large surface protein, p45, spans the preS1, preS2 and S domains.

1.1.2.4 Class IV: The 0.7 kb transcript

A polyadenylated transcript of about 0.7 kb serves as template for the synthesis of the X protein. This transcript has not been detected in HBV-infected liver nor in hepatoma cells transfected with replication-competent viral DNA genomes, and is thought to be drastically repressed under conditions which support viral replication (Kaneko and Miller, 1988). Two

hundred basepairs upstream of the X transcription initiation site is the X promoter which contains the HBV enhancer element, Enhancer I (ENHI). This enhancer is under the control of regulatory factor X1 (RFX1, which binds to the enhanced factor C [EF-C] site) and HNF1 as well as a liver-specific regulatory element, LSR. The LSR element (nucleotide 1115-1149) contains two binding sites (2C/TGT3b and GB/RARE) to which HNF3 and members of steroid hormone receptor superfamily (such as HNF4, PPAR [peronisome proliferator-activated receptor] and RXRα [retinoid X receptor α]) bind *in vitro* and which regulate ENHI in a synergistic fashion (Appendix A, Fukai *et al.*, 1997).

1.1.2.5 Spliced transcripts

Chen *et al.* (1989) identified two spliced transcripts in chronic patients. The P, or *pol*, ORF has three functional domains (the terminal protein [TP], the reverse transcriptase [RT] or RT/*pol* domain, and the putative protease, ribonuclease H [RNAse H]) and includes a hypervariable region, or spacer region, at its amino terminus which is removed by the first splicing event. The second splicing event deletes 282 bases in the central region of the HBcAg and brings the TP domain closer to the rest of *pol*. However, since mutational inactivation of splice donor or acceptor sites does not result in a detectable loss of competence for viral replication, the functional significance of these splicing events is unclear. Another hepadnaviral spliced transcript has been found to be essential for replication. This major spliced mRNA, termed spl L-RNA, results from splicing of the RNA pregenome and encodes for the LHBs (Obert *et al.*, 1996). A spl L-RNA negative mutant was unable to replicate in competition with low amounts of wild type virus in infections of primary duck hepatocytes, indicating that spl L-RNA is required for a productive infection. Two mRNAs with differing 5' ends for the translation of one protein would confer a distinct advantage to the virus *in vivo*. No evidence for a similar spliced transcript in HBV has yet been demonstrated.

1.1.2.6 Regulation of HBV RNA export

The steady-state ratio of transcripts observed *in vivo* is S>C>>S1>=X. HBV has a post-transcriptional regulatory element (PRE) consisting of two *cis* RNA sub-elements which are thought to interact with cellular proteins, which directly export the unspliced HBV RNAs. These RNA sub-elements are approximately 500 bp long (nucleotide1203-1515) and partially overlap the X ORF. The two sub-elements function synergistically to control the

efficiency of RNA export in a similar manner to the *rev* binding site of the human immunodeficiency virus (HIV; Patzel *et al.*, 1997).

1.1.2.7 3' end formation

All hepadnaviral RNAs terminate at the common polyadenylation site 20 nucleotides downstream of the conserved hexanucleotide TATAAA which starts at nucleotide 1916 in HBV. The polyadenylation signal of the hepadnaviruses (UAUAAA) varies from the highly conserved sequence motif (AAUAAA) which is important for the processing of newly synthesized mRNA and so is intrinsically weak. Over-length transcripts are generated when this signal is recognized only after the transcription complex passes through the signal a second time.

Huang and Summers (1994) describe two *cis*-acting regions within the DHBV genome involved in regulating pregenome transcription. *Pet*, the positive effector of transcription, is present in the 5' non-translated region of the pregenome and its deletion results in a great reduction in transcription from the pregenome promoter. However, a second region, ***net***, the negative effector of transcription, is located approximately 200 to 900 nucleotides downstream from the polyadenylation signal. Deletion of both regions restores transcription of the pregenome to wild type levels, thus it is thought that *net* blocks pregenome expression and *pet* overcomes this block. These data have been interpreted to indicate a role for *net* in termination of transcription and hence 3' end formation of viral mRNA. Since ***net*** is located between the core and the preS1 promoter, its effect would result in the premature termination of the core transcripts, but the correct termination of all other transcripts. The purpose of *pet* is then postulated to be overriding the effect of ***net*** in a distance-dependent manner, so that mature core transcripts of slightly greater than genome length would be produced.

1.1.3 The structure and function of the 7 HBV Proteins

Seven HBV proteins have been characterized and, since the virus uses all three reading frames, there is overlap (surface/P, X/P, X/core and core/P) between the four viral open reading frames i.e. the same nucleotides code, in different reading frames, for more than one of the viral proteins.

1.1.3.1 The core (*gag*) gene products

Two proteins are generated from the core coding region: The soluble and secreted e antigen, HBeAg, and the nucleocapsid protein, HBcAg. Although initially reported to be well conserved, this region is now known to undergo evolution within patients (Uchida *et al.*, 1994) and considerable mutation in chronic active carriers (Carman *et al.*, 1997). During infection antibodies are developed against both HBcAg and HBeAg. Some cross-reactivity has been observed between these two antibodies and complementary specificities, "e1" and "e2", have been identified in antisera directed against HBeAg. "e1" reacts with the assembled core protein while "e2" only reacts with HBeAg. "e1" is found in the exposed region at the capsid surface and is accessible in intact cores, whereas "e2" involves portions of the negative strand and a loop region at the five-fold interaction site and is buried in the assembled core (Figure 2).

1.1.3.1.1 The core antigen, HBcAg

The HBcAg is the major product of the core gene and the main structural component of the viral nucleocapsid. The core antigen is initiated from the second ATG codon at the start of the core region (nucleotide1901, Appendix A). Core particles are assembled from dimers of a single 185-residue viral capsid or core protein (p21.5) which possesses two distinct domains. Residues 1 to 144 form a minimal capsid assembly domain and the arginine-rich, carboxy-terminal residues, 150 to 185, form a protamine-like domain that mediates nucleic acid binding (Seifer and Standring, 1994). The HBcAg has been shown *in vitro* to have a suppressive effect on the human interferon promoter (Twu *et al.*, 1988; Whitten *et al.*, 1991) which could explain the inability of chronic patients to produce interferon.

Core protein has been found both in the cytoplasm and in association with the nuclear membrane of HBV-infected hepatocytes. Maturation-controlled phosphorylation is required during the life cycle of HBV for the transport of the core particles to the nuclear pores where the DNA genome is released into the nucleoplasm (Yeh *et al.*, 1993). Binding of core to the nuclear membrane has been reported but the nuclear localization signal is not thought to be sufficient for translocation of the core through the nuclear pore complex (Guidotti *et al.*, 1994; Bock *et al.*, 1996).

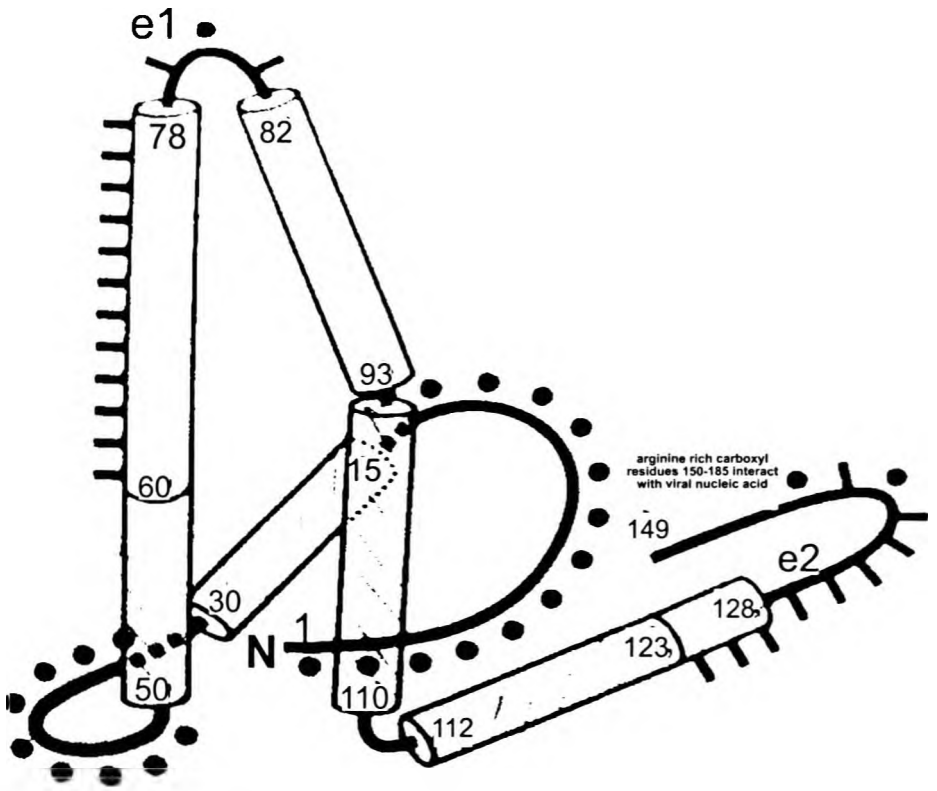


Figure 2. Model of the core antigen (HBcAg) showing core protein domains as described by Koschel *et al.* (1999). Alpha helical residues are drawn as cylinders. Numbers indicate amino acid residues. Regions where insertions or deletions were compatible with capsid formation are marked by dots - these correspond to non-helical domains. Areas where insertions or deletions allow stable HBcAg or HBeAg expression but no capsid formation are indicated by bars. Picture adapted from Bottcher *et al.* (1998)

Although antibodies to the HBV core antigen do not neutralize the virus, nucleocapsid specific cellular immune responses are important in virus elimination. Strong T-cell dependent and independent responses are elicited in 100 percent of patients (Milich *et al.*, 1987). Protection attained by vaccination with core is due to its ability to induce a strong cellular response.

1.1.3.1.2 The e antigen, HBeAg

A second core gene product, the e antigen or HBeAg, is generated from a 25 kilodalton (kd) precursor protein when transcription is initiated from the first ATG codon of the precore region (nucleotide1814, Appendix A). The first maturation step is the removal of the amino-

terminal signal sequence. The arginine-rich domain located at the carboxy-terminus is then removed in a multi-step process. The precore region targets the antigen to the endoplasmic reticulum (ER) and results in its secretion (Ou *et al.*, 1986). HBeAg is also expressed on the cytoplasmic membrane of infected cells. It crosses the placenta and acts as a tolerogen in the neonate guaranteeing viral persistence after perinatal infection. It is also hypothesised that T-cell tolerance to HBeAg plays a role in chronicity and that HBeAg-specific T-cells emerging from the tolerant state mediate liver injury and HBV clearance in chronic hepatitis B infection (Milich *et al.*, 1993).

The p22 non-secreted precore precursor protein has been shown to be a potent natural inhibitor of HBV replication when transfected into hepatocyte derived lines (Scaglioni *et al.*, 1997).

1.1.3.2 The surface (*env*) gene products

In infected cells HBV induces overproduction of its three surface proteins all of which have the same terminal sequences. They are produced from the same ORF having a common stop codon and three separate in-phase start codons spaced at intervals of 108 (or 119, depending on genotype) and 55 codons, respectively. The preS1 region extends from the first amino acid and together with the preS2 and S regions encodes LHBs. The preS2 region begins at amino acid 109 (or 120) and together with the S region encodes MHBs. Finally, the S region begins at amino acid 164 (or 175) and encodes the 226 amino acids of SHBs or HBsAg.

Characteristic of the hepadna family of viruses are the abundant non-DNA containing particles present in serum with the virion particles. These non-infectious particles have two forms: the circular 20 nanometre (nm) particles and the filamentous forms (with a mean diameter of 20 nm). The virion outer coat contains equimolar amounts of all surface antigen polypeptides (Tiollais *et al.*, 1981; Hollinger, 1990). All three morphological forms of HBsAg consistently contained six protein bands (P24, GP27, GP33, GP36, P39 and GP42) and the relative staining intensities of P24, GP27, GP33 and GP36 were similar in all three forms. However, the intensity of the L protien, P39 and GP42, differed significantly among the three forms. The composition of the filaments is identical to the complete virions which contain 300-400 HBsAg molecules and 40-80 middle and large protein molecules but the 20 nm particles contain only one or two GP42/P39 molecules (Heermann *et al.*, 1984).

1.1.3.2.1 The small (major) surface antigen, SHBs (HBsAg)

SHBs or HBsAg, p24S, or its glycosylated form, gp27S, is a hydrophobic membrane-spanning polypeptide with the ability to bud as a 20 nm particle from the membrane of the ER to the ER lumen. The primary structure of the SHBs can be divided into three hydrophobic regions (amino acids 8-28, 80-98, and 161-226, respectively) separated by two hydrophilic regions (amino acids 29-79, and 99-160 respectively; Mangold *et al.*, 1997). The first hydrophobic region forms the first membrane spanning helix of S and is called signal I since it is able to insert into the ER membrane and mediates the translocation of upstream parts of the protein (Figure 3). The hydrophilic residues 29-79 are intracellular, located within the HBsAg particles, and cannot be recognized by B cells. Residues 80-98 are also able to insert into the ER membrane and are called signal II, or the second membrane spanning helix (Stirk *et al.*, 1992). Signal II mediates the translocation of the downstream sequences of S. Both Signal I and II are candidates for membrane anchoring (Berting *et al.*, 1995) and are important for virion assembly. The second hydrophilic domain is external and residues 139-147 form the HBV-specific "a" determinant of the surface antigen, common to all serological subtypes. There are 8 cysteines in the external region of SHBs and intra-chain disulphide bridges stabilise the structure (Mangold *et al.*, 1995). Asparagine 146 (^{146}Asn) is the site of glycosylation. The amino residues 1-8 and the carboxy-terminus have also been reported to be external.

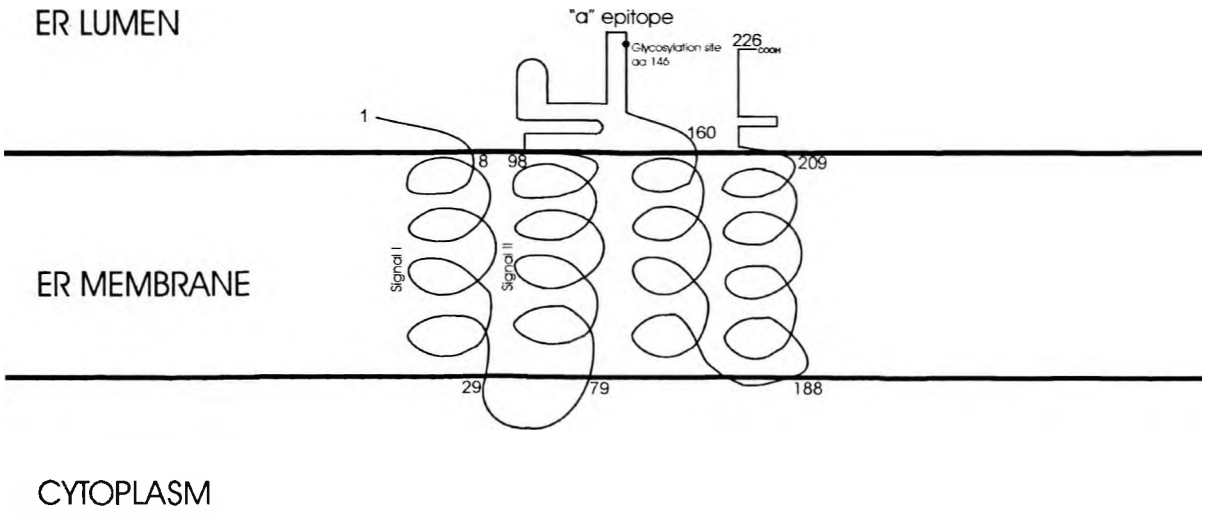


Figure 3. Model of HBsAg in the membrane showing the "a" epitope and the glycosylation site (adapted from the proposed model for HBsAg of Stirk *et al.* [1992])

Secondary structure predictions support the presence of four helices. Helices III and IV form the third hydrophobic region and can be orientated in such a way that interaction between hydrophilic areas are stabilized by hydrogen bonds, whereas the other regions are stabilized by hydrophobic interactions. SHBs is able to form dimers and the amphipathic helices (I, III and IV) can be arranged to form a hypothetical membrane channel approximately 1.5 nm in diameter which would form a pore through which small particles and molecules could pass through the hydrophobic envelope (Stirk *et al.*, 1992).

1.1.3.2.2 The middle surface protein, MHBs

MHBs differs from SHBs by the addition of 55 amino acids at the N-terminal which are coded for by the preS2 domain. Its biosynthesis is very similar to the synthesis of SHBs. The preS2 domain is cotranslationally translocated into the ER lumen by signal I in its downstream S domain (Eble *et al.*, 1990). PreS2 becomes N-glycosylated at ^{4}Asn of MHBs to form a 33 kd protein. Additional N-glycosylation in the S domain results in a 36 kd protein. The MHBs protein also forms cysteine-linked dimers although there is no cysteine residue in the preS2 domain.

A carboxy-terminally truncated form of the middle surface protein functions as a transcriptional activator. The truncated large surface protein also showed this effect but the preS1 region seemed to be dispensable. Deletion or mutation analysis of preS2/S sequences revealed that gene truncation had to occur within a defined region of the S gene between S amino acids 22 and 139 (Caselmann, 1996) but later studies showed that even the presence of preS2 amino acids 1-54 may be sufficient to exert the *trans*-activating effect (Hildt *et al.*, 1996).

1.1.3.2.3 The large surface protein, LHBs

LHBs carries (relative to MHBS) the additional N-terminal 119 amino acids of the preS1 domain. The preS domains of LHBs stay at the cytosolic side of the ER during synthesis of the protein (Prange and Streek, 1995) and do not contain a functional signal sequence. LHBs is anchored in the membrane by signal II which functions at an internal position and directs translocation of downstream sequences. The luminal loop in the S domain of LHBs is partially N-glycosylated like MHBs and SHBs protein. ^{4}Asn in preS2 is not glycosylated in

LHBs. The large protein forms mixed disulphide-linked dimers with MHBs and SHBs proteins. The preS1 sequence contains no cysteine residue. It is myristylated at glycine 2 (^{2}Gly) of the preS1 although it is not known whether this fatty acid moiety interacts with the membrane or whether it is embedded in a hydrophobic region of the protein. The large protein adopts two different transmembrane topologies due to a post-translational switch of the folding and forms two types of molecule with different functions (Bruss *et al.*, 1994). LHBs molecules with the preS1 domain exposed [e-preS, Figure 4(a)] on the viral particle surface are thought to be ligands for the putative virus receptor. The specificity of this ligand-receptor interaction determines species specificity and liver tropism of the virus. LHBs chains with internal preS1 domains [i-preS, Figure 4(b)] are required for virion morphogenesis and act like a matrix protein to mediate contact to the nucleocapsid.

The LHBs protein is phosphorylated and the major site of phosphorylation is serine 118 (^{118}Ser). The functional significance of phosphorylation appears to be related to the post-translational shift in topology of the preS domain. Overexpression of this form of LHBs protein inhibits viral particle release in a dose-dependent manner. In the transgenic mouse

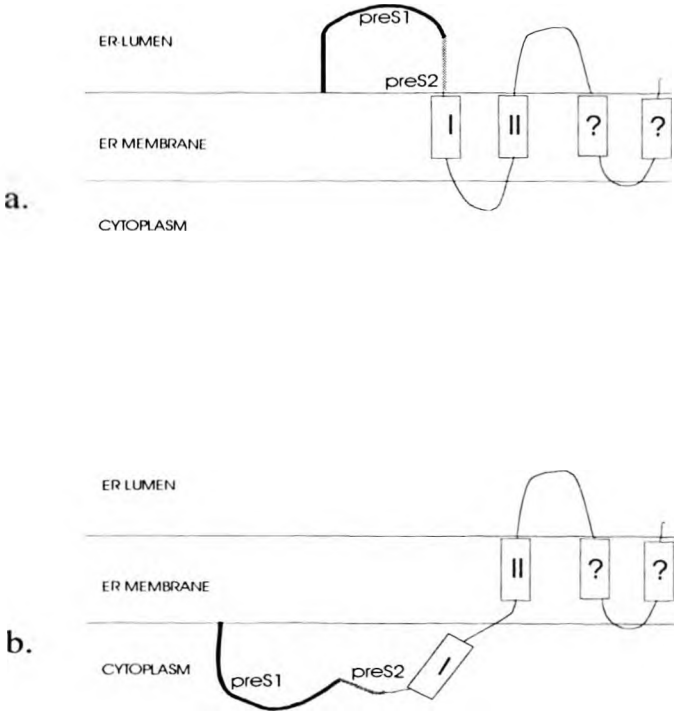


Figure 4. The two different trans-membrane topologies adopted by the large surface protein (a) e-preS in which the preS1 domain is exposed on the surface and (b) i-preS with internal preS1 domain (adapted from Bruss *et al.* [1994])

model the retention of surface proteins is cytotoxic and can eventually result in cancer development (Chisari *et al.*, 1989). The histology of the mouse cells resembled the ground-glass hepatocytes observed in humans. Mutation of the acceptor glycine (amino acid 2 or 19 depending on genotype) abolishes the retention function (Kuroki *et al.*, 1989).

Anti-preS1 is usually the first immune response followed by the anti-HBs response which, in most individuals, is immunogenic and protective against subsequent challenge (Hollinger, 1990).

1.1.3.3 The P (*pol*) gene

Although the actual mechanism of P protein translation is unclear, it has been established that P is synthesized separately from the core protein. The polymerase ORF spans approximately two-thirds of the virus genome and overlaps the ORFs of the three other virus genes (core, surface, and X). The ORF codes for three functional domains and the minimum and functional boundaries of the terminal protein and the RT/*pol* were recently determined (Lanford *et al.*, 1999). Miller (1988) predicted the probable boundaries from conserved regions of the sequence, and these agree quite well with those which Lanford determined using priming and binding assays (see Figure 5 and legend). The first functional domain is the terminal protein (TP) which Miller predicted to be between amino acids 50-

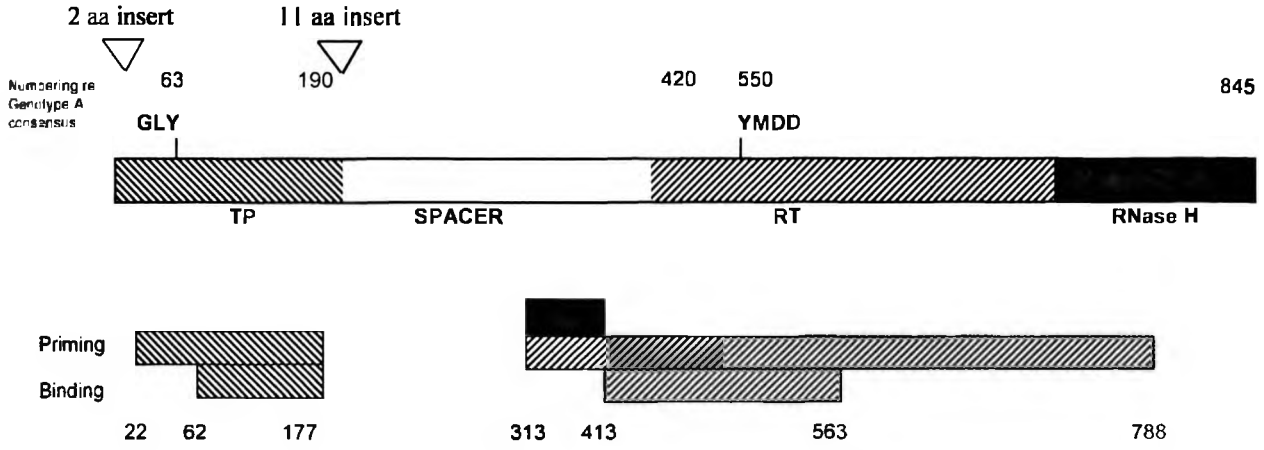


Figure 5. The functional domains of *pol* shown in the context of the full-length HBV polymerase. The boundaries are shown for the minimal TP and RT domains functional in a *trans*-complementation assay for protein-primed reverse transcription (priming) and the minimal TP and RT domains capable of protein-protein interaction in a binding assay (binding). The inhibitory domain that inhibits *trans*-complementation when spliced to the TP domain is shown as a black box above the minimal RT domains. Adapted from Lanford *et al.* (1999)

190 of the *pol* protein. Also called the primer region, TP was found by Lanford to have a minimal domain mapped to amino acids 22 to 177, however, maximum activity required amino acids 1 to 212. TP binds covalently with the first nucleotide of the minus-strand DNA at amino acid 63 (Lanford *et al.*, 1997) preventing phosphorylation of the strand. Expression of the terminal protein region of HBV inhibits cellular responses to interferons α and β (Foster *et al.*, 1991). This domain is separated from the reverse transcriptase (or RT/*pol*) domain by a hypervariable domain called the spacer region predicted by Miller to include polymerase amino acids 190-420. Lanford found the minimal RT domain spanned residues 313 to 788. A portion of the spacer was found to be required for both TP and RT functions. The RT/*pol* domain forms a stable complex with TP for nucleotide priming and reverse transcription. The *pol* has DNA- and RNA-dependent DNA polymerase activities. The last domain of P codes an RNase H activity. Miller predicts this domain to begin at amino acid 675 and the stop codon of P is at residue 845. This domain is separable from the polymerase activity (Oberhaus *et al.*, 1995) although Lanford found that priming required the RNAse H domain. The RNAase H domain is thought to stabilize the TP-RT interactions with the template or cellular proteins rather than playing a direct role in priming.

For consistency, numbering throughout is based on (and converted to, where necessary) the genotype A consensus since this is the longest genome of HBV. Genotype A, only, has a two amino acid insert at amino acids 17 and 18 of P and a 33 bp insertion (present in all except genotype D) at amino acids 183 to 194 at the start of the spacer region.

1.1.3.4 The X gene product, HBx

The smallest open reading frame (465 bases) codes the X protein (Kay *et al.*, 1985) which has a molecular mass of 17,000. Antibodies to X have been detected in the sera of infected patients and the X antigen, HBxAg, has been found in the liver of infected individuals and in human hepatoma cells (Moriarty *et al.*, 1985; Stemler *et al.*, 1990). X has been variously localised in either the cytoplasm or the nucleus, or both (Rossner, 1992). Although X is well conserved in the mammalian hepadnaviruses, the avian hepadnaviruses lack this ORF (Koike, 1995). Since the risk of developing hepatocellular carcinoma (HCC) is greatly enhanced in chronic infection with the mammalian (Snyder *et al.*, 1982; Marion *et al.*; 1986; Szmuness, 1978; Beasley *et al.*, 1981) but not the avian hepadnaviruses, X has long been suspected of playing a role in liver carcinogenesis (Koike *et al.*, 1989). X was found to

induce tumours in some transgenic mouse strains (Hohne *et al.*, 1990; Kim *et al.*, 1991) and to make other strains more sensitive to hepatocarcinogens (Slagle *et al.*, 1996). It has been shown to be important in the establishment of infection in woodchucks (Chen *et al.*, 1993) and to be a serine protease inhibitor (Takada and Koike, 1994) but it is most widely reported to function as a transcriptional *trans*-activator.

Within the X ORF are the direct repeat sequences (DR1 and DR2), which are vital for

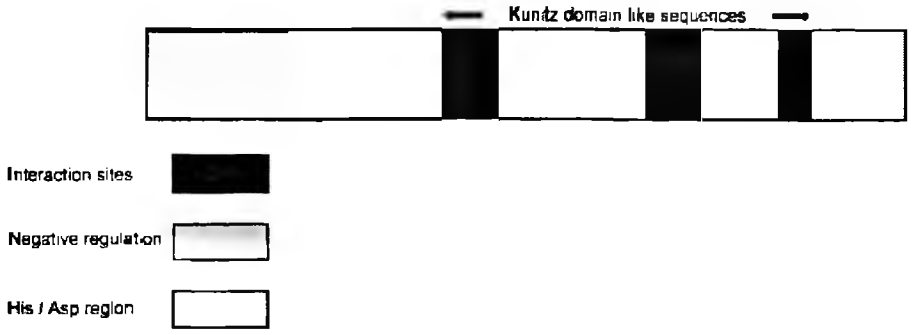


Figure 6. Schematic representation of the characteristic regions in the HBV X protein

replication, as well as most of the core promoter region and its adjoining enhancer region, enhancer II (ENH II; Yuh *et al.*, 1992). The X protein has amino acid sequence homologous to the functionally essential region of Kunitz-type serine protease inhibitors and within this "Kunitz domain" three separate sites (amino acid 65-72, 105-115 and 131-142) have been found to interact with cellular proteins. Futhermore, mutations in these regions have been found to abolish the *trans*-activating ability of X (Takada and Koike, 1994). The N-terminal region (amino acid 2-30) is not essential for the *trans*-activation function itself but is involved in negative regulation of *trans*-activation (Takada and Koike, 1990) and a histidine/aspartic acid region (His/Asp region amino acid 46-52) was found to be important for *trans*-activation under the conditions used (Arii *et al.*, 1992; Figure 6).

Because the X gene product exerts pleiotropic molecular effects at multiple levels it has been compared to other viral *trans*-activators e.g. Tax (Yen, 1996). It profoundly influences cellular gene expression upregulating transcription from a wide variety of cellular and viral promoters and enhancers including those controlling HBV gene expression and replication (Colgrove *et al.*, 1989). At the same time HBx does not bind double-stranded DNA directly and must therefore effect its *trans*-activation function by interacting with cellular proteins

(Takada and Koike, 1994).

X mimics the activity of tumour promoters and phorbol esters by activating signal transduction pathways which in turn affect many transcription factors important for regulation of cell growth (Gilman, 1993). These include the signalling cascades of the *Ras-Raf* MAP-kinase pathway (Benn and Schneider, 1994) and the protein kinase C-dependent signalling pathway (PKC; Kekul *et al.*, 1992; 1993). PKC in turn activates reporter genes under the control of signal-responsive elements such as activating protein 1 (AP-1) cyclic adenosine 5'-monophosphate responsive element (CRE) and the κB elements. NFκB is involved in the transient expression of genes important for cell proliferation and acute inflammatory responses. Host genes shown to be activated in cells transfected with the X gene include interleukin-8 (IL-8; Mahe *et al.*, 1991), the proto-oncogenes c-*myc* (Benn *et al.*, 1996) and c-*jun* (Twu *et al.*, 1993), tumour necrosis factor α (TNF-α; Lara-Pezzi et al., 1998), and the human leucocyte antigen DR molecule (HLA-DR; Hu *et al.*, 1990). Although transcription factors respond to signal transduction, there is also evidence for direct interaction of X with the transcription machinery at the promoter level (Koshy and Meyer, 1992). For example, X forms heterodimers with the RPB5 subunit of RNA polymerase II (Lin *et al.*, 1997) and acts directly with the CREB/ATF transcription factor family (Maguire *et al.*, 1991) and the TATA-binding protein (TBP; Qadri *et al.*, 1995; Figure 7). X also acts directly on the NFκB promoter and has been shown to *trans*-activate the promoter of the κB-like enhancer sequences in the long terminal repeat (LTR) of HIV (Twu *et al.*, 1989) and the HBx promoter itself. Thus it is not surprising that X has the capacity to deregulate cell cycle control shortening checkpoint intervals by activation of *Ras* and other unknown effectors (Benn and Schneider, 1995).

In addition X has been found to activate PKC indirectly by causing production of the membrane-associated lipid activator diacylglycerol, DAG, which results in the rapid translocation of PKC from the cytosol to the membrane (Luber *et al.*, 1993). By triggering this phospholipase, X mimics an activity of growth factors and hormones (Gilman, 1993). Three independent assays have shown that X compromises DNA repair by binding to the UV-damaged DNA binding protein (UV-DDB or XAP-1) predisposing cells to apoptotic killing (Lee *et al.*, 1995). X also binds and inactivates negative growth regulatory molecules such as p53 (Wang *et al.*, 1994). Together these two properties may promote genomic

instability (Feitelson and Duan, 1997). X also stimulates positive growth regulators such as insulin-like growth factor II (Su *et al.*, 1994) and stimulates the expression of the insulin-like growth factor I receptor (Kim *et al.*, 1994). This could set up an autocrine loop which would enhance cell growth (Feitelson and Duan, 1997). The many mutually interfering but distinct activities of X which have been demonstrated *in vivo* and *in vitro* (Koike, 1995) could quite easily result in the selection of genetically unstable cells in which unrepaired, transforming mutations occur and as such represent a likely molecular mechanism by which HBx contributes to viral carcinogenesis (Benn and Schneider, 1995), as suggested from the data from the avian and transgenic mouse models.

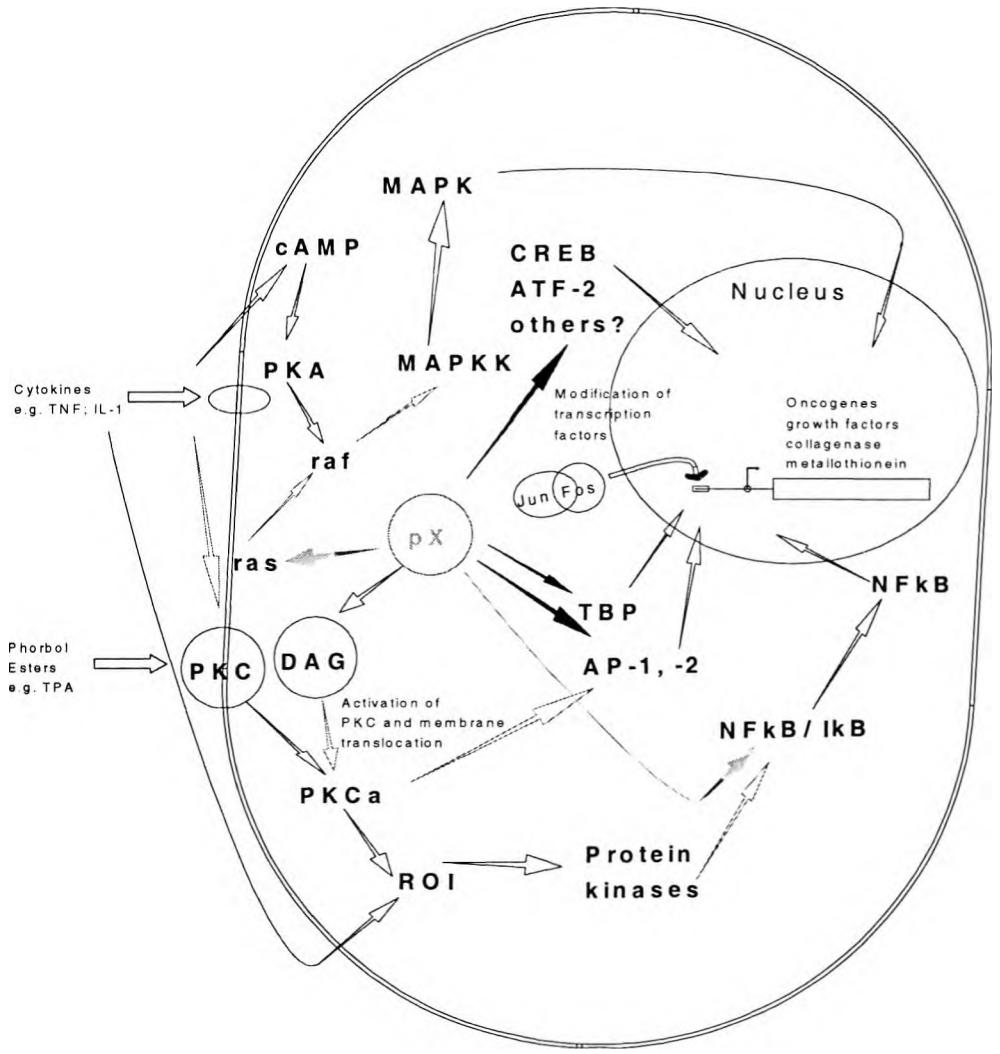


Figure 7. Some of the pathways which may be altered by HBx as described in the text (adapted from Feitelson and Duan, 1997)

The X protein may up– or down-regulate the pathways shown with the black arrows by directly binding to one or more critical components of the pathways such as the TATA-binding protein (TBP), activating transcription factor-2 (ATF-2), cAMP response element-binding protein (CREB) or activating protein-1 and –2 (AP-1 and –2). Alternatively, or in addition, X may indirectly stimulate the Ras and/or NFκB signal transduction pathways (grey arrows).

TPA (12-o-tetradecanol phorbol-13 acetate); cAMP (cyclic adenosine 5'-monophosphate); PKA/C (protein kinase A/C); DAG (diacylglycerol); ROI (reactive oxygen intermediates); IκB (NFκB inhibitor); MAPK (microtubule associated protein kinase); MAPKK (microtubule associated protein kinase kinase)

1.2. The immunopathogenesis of HBV

HBV is not directly cytopathic for the infected hepatocyte. Both the humoral and cellular limbs of the immune response are required for viral clearance together with a non-cytolytic intracellular viral inactivation by cytokines released by virus-activated lymphomononuclear cells (Chisari and Ferrari, 1995). The host immune response is not only responsible for viral clearance but also for disease pathogenesis during HBV infection. Most of our knowledge of viral pathogenesis has come from experimental studies using the HBV transgenic mouse model since traditional experimental approaches cannot be applied due to the limited host range and organ tropism of HBV, together with the lack of an *in vitro* culture system to propagate the virus.

1.2.1 The life cycle of HBV

HBV productively infects only hepatocytes of humans and some higher primates and the precise mode of cell binding and entry is not yet known.

1.2.1.1 Viral binding and cell entry

Since N-terminal mutations of the S ORF do not affect virion formation but do destroy infectivity, the preS has long been known to be involved in the complex series of events which begin with cellular receptor binding and culminate in the productive infection of a cell by hepadnaviruses (step 1 of Figure 8; Lenhoff and Summers, 1994). However, since target cell infection was found to be dependent on LHBs but not MHBs (Fernholz *et al.*, 1993) the search has been narrowed to the preS1 region. The preS2 domain in MHBs has several non-essential liver attachment sites including the polymerized albumen receptor domain (Imai *et al.*, 1979) and the N-linked glycan at the amino end of the preS2 domain (Gerlich *et al.*, 1993). It has also been shown to bind the extracellular matrix fibronectin of human liver sinusoids (Budkowska *et al.*, 1995) and the transferrin receptor (Franco *et al.*, 1992). There have been numerous proposals for the putative, still unidentified, liver-specific preS1-receptor (de Meyer *et al.*, 1997). These include the IgA receptor (Pontisso *et al.*, 1992), IL-6 (Neurath *et al.*, 1992), the asialoglycoprotein receptor (Triechel *et al.*, 1997), human liver glyceraldehyde-3-phosphate-dehydrogenase (GAPD, Petit *et al.*, 1992) and the liver-membrane located human annexin V (hAV, previously called Endonexin II; Hertogs *et al.*, 1993). More recently several groups identified a putative component of the cell receptor

complex which takes part in the early steps of DHBV infection. This was found to be a 180 kd carboxypeptidase D (gp 180; Urban *et al.*, 1999).

HBsAg alone without the preS region was also found to bind to hepatocytes via apolipoprotein H (Mehdi *et al.*, 1994; 1996) and hAV (de Meyer *et al.*, 2000). hAV is a Ca^{2+}-dependent phospholipid-binding protein and de Meyer *et al.* (2000) claim further that species-specific susceptibility to HBV infection and replication in hepatocytes is associated with hAV expression.

1.2.1.2 cccDNA synthesis and integration

After cell entry, the relaxed circular (RC) partially double-stranded viral genome, within the virus core, is transported to the nuclear membrane without processing (step 2 of Figure 8). Here it is uncoated and enters the nucleus. Within the nucleus the viral DNA is converted to covalently-closed, circular, supercoiled DNA molecules (cccDNA; step 3 of Figure 8). cccDNA is the major form of viral DNA found in the nucleus of hepadnavirus-infected hepatocytes. The conversion process requires the removal of the covalently bound protein, one copy of the terminally redundant segment from the 5' end of the minus strand as well as the RNA primer from the 5' end of the plus strand. The plus strand is then completed and both strands are ligated (Mason *et al.*, 1983). The exact mechanism of conversion is unknown but it is thought to involve cellular enzymes (Kock and Schlicht, 1993). Eukaryotic topoisomerase I (top I; Wang and Rogler, 1991) is capable of processing hepadnavirus DNA towards either linear or circular structures *in vitro* and involves non-homologous recombination (Pourquier *et al.*, 1999). These competing reactions could regulate accumulation of circular (replication competent) versus linear (integration precursor) DNA molecules in the nucleus although integration of HBV DNA into the host genome does not occur during the normal course of replication, as it does with retroviruses (Lau and Wright, 1993).

The cccDNA appears prior to viral-specific RNA and is therefore thought to be the template for the transcription of viral RNAs. Immediately post-infection there is a rapid 50-fold amplification of cccDNA which reaches peak levels by day four post-infection. Tuttleman *et al.* (1986) found that cccDNA did not replicate primarily via a semi-conservative mechanism but was formed, almost entirely, via an intracellular cycle of replication similar to the normal reverse transcription pathway of DNA synthesis (steps 4-7 of Figure 8). Thus

The virus is uncoated and enters the cell (step 1), the relaxed circular genome (RC-DNA) is then transported within the core particles to the nuclear membrane (step 2) where the capsid is removed and the RC-DNA enters the nucleus. RC-DNA is then repaired to form covalently closed circular DNA (step 3) from which the pregenome and various mRNAs are transcribed (step 4). The binding of polymerase to the RNA encapsidation signal, ε, signals the start of both replication and nucleocapsid assembly (step 5) and this is followed by plus strand synthesis (step 6). At this point the mature genome within the core particle can be transported back to the nucleus (step 7) or, alternatively, acquire a virion envelope (step 10) and be exported into the bloodstream (step 11). Step 8 represents the translation of the viral mRNA to viral gene products and the grey arrows show that these proteins provide the capsid coat (also shown in grey) for the pregenome as well as the surface antigen necessary for viral assembly at step 10. The relative abundance of viral proteins is thought to control the fate of the mature genome within the core particle. Too much viral protein blocks the feedback mechanism of step 7 and this is shown by the grey arrow and grey cross on the feedback arrow of step 7..

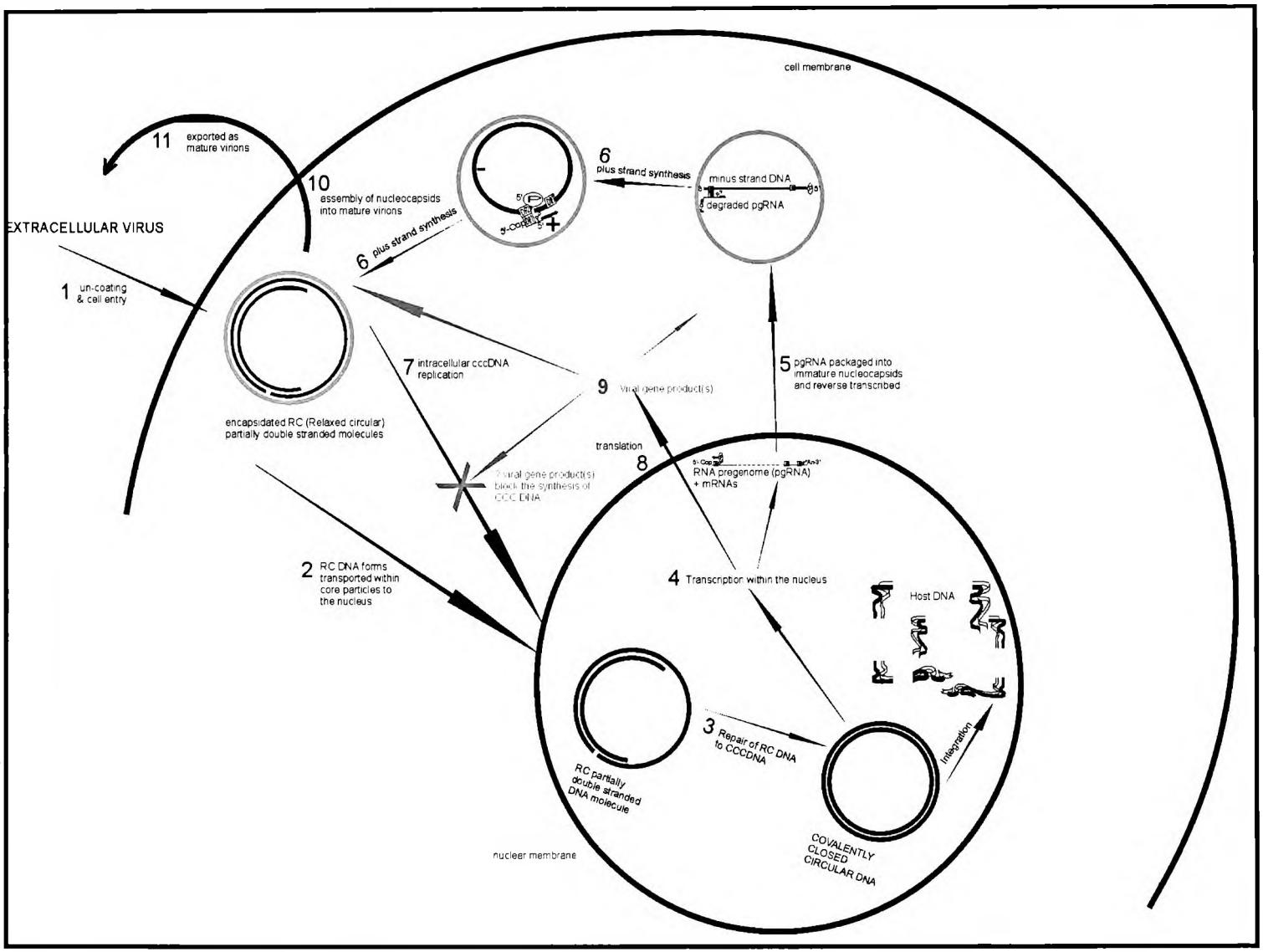


Figure 8. The life cycle of HBV

there are three possible fates for the mature genome within the core particle: It can acquire a virion envelope and be exported into the bloodstream (step 10 and 11 of Figure 8), remain in the cytosol and be degraded, or be transported back to the nucleus to maintain persistent infection of the cell (step 7 of Figure 8). The production of viral gene products (step 8 of Figure 8) is thought to influence the fate of the core particle by negative control of the intracellular loop (step 9 of Figure 8) which prevents further conversion of RC DNA into cccDNA and results in the exclusive assembly of cytoplasmic nucleocapsids into mature virus.

1.2.1.3 Viral replication

The first stage in the replication of HBV is unique and involves reverse transcription of the greater than genome size RNA pregenome (Summers and Mason, 1982b; Figure 9 a-d). Priming of this reaction is assisted by the heat shock protein, Hsp90, which interacts with the viral reverse transcriptase domain of P and facilitates the formation of a ribonucleoprotein (RNP) complex between the polymerase and the pregenome RNA (pgRNA; Hu and Seeger, 1996). Binding is at the RNA encapsidation signal, ε, at the 5' end of the pgRNA (nucleotide 3190; Figure 9a). This binding simultaneously initiates replication and nucleocapsid assembly in the cytosol (step 5 of Figure 8). After synthesis of a few nucleotides, copied from the bulge of ε, the covalent complex is translocated to nucleotide 8 in DR1* at the 3' end of the pgRNA (Figure 9b) and primer extension continues from this point (Wang and Seeger, 1993). As the minus strand is synthesized, the RNA pregenome is concomitantly digested (Figure 9c and d) resulting in a DNA minus strand terminally redundant by five to eight nucleotides and still attached to the short oligoribonucleotide derived from the 5' end of the pgRNA (Nassal and Schaller, 1996). After first being transposed to DR2 at the 3' end of the DNA minus strand, this oligonucleotide serves as the primer for DNA plus strand synthesis (Figure 9e). Plus strand synthesis can only proceed up to the 3' end of the DNA minus strand and then a template switch has to take place before further elongation can occur. This template switch is facilitated by the short terminal redundancy of the DNA minus strand and results in the circularisation of the minus strand DNA (Figure 9f). The elongating DNA plus strand then stabilizes this circular intermediate.

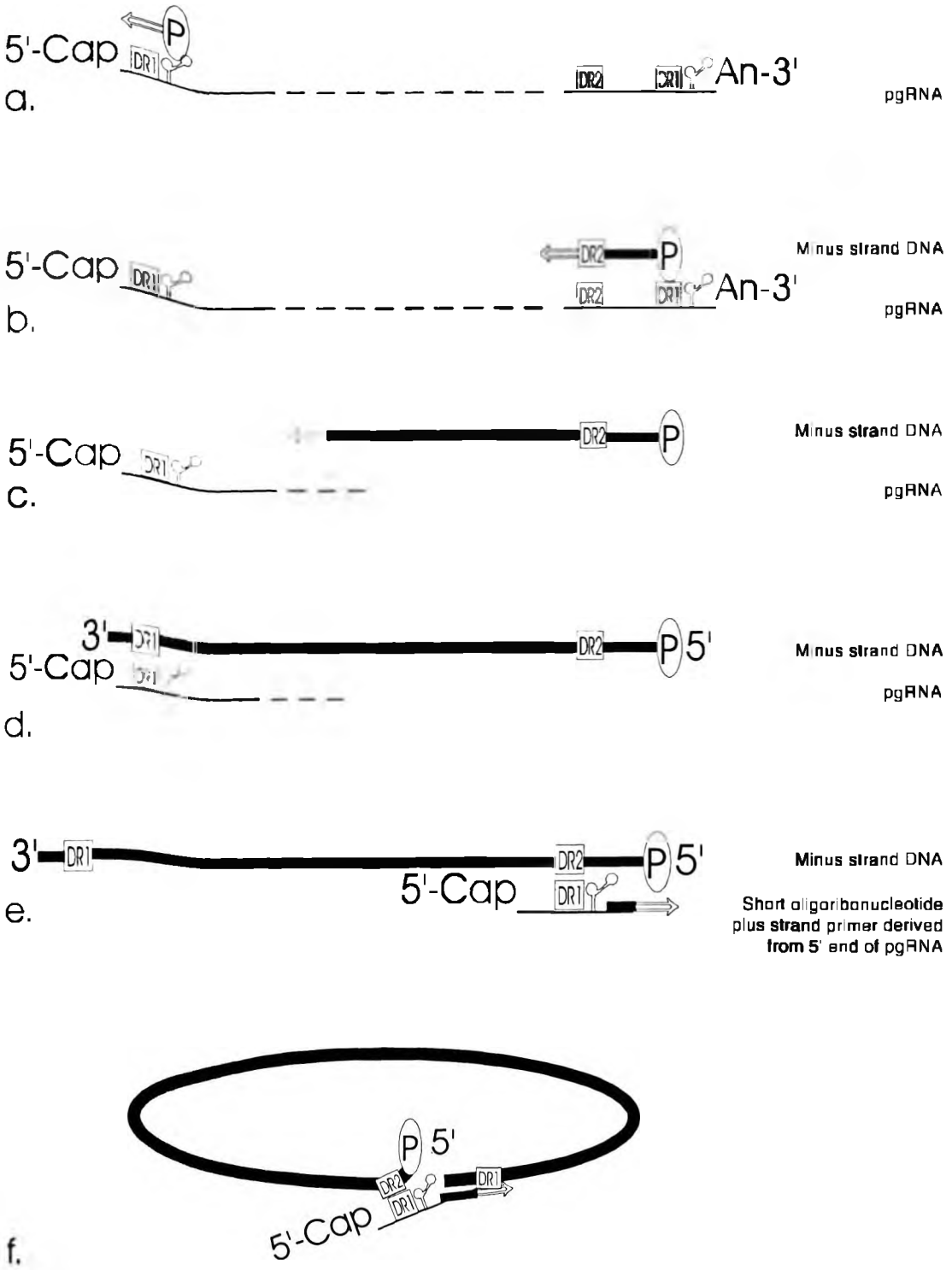


Figure 9. Replication strategy of HBV (adapted from Tavis *et al.* [1994])

(a) Replication begins with the binding of viral polymerase to the RNA encapsidation signal, ε, at the 5' end of the pgRNA (b) After synthesis of a few nucleotides, copied from the bulge of ε, the covalent complex is translocated to nucleotide 8 in DR1* at the 3' end of the pgRNA and primer extension continues from this point. (c) & (d) As the minus strand is synthesized, the pgRNA is concomitantly digested resulting in a terminally redundant DNA minus strand attached to a short oligoribonucleotide derived from the 5' end of pgRNA. (e) After first being transposed to DR2 at the 3' end of the DNA minus strand, this oligonucleotide serves as the primer for DNA plus strand synthesis which can only proceed up to the 3' end of the DNA minus strand when a template switch is necessary. (f) The template switch is facilitated by the short terminal redundancy of the DNA minus strand and results in the circularization of the minus strand DNA.

1.2.1.4 Packaging and export

All three surface antigen proteins are synthesized at the membrane of the ER and homo- and hetero-dimers of surface antigen form in the pre-Golgi compartment and eventually bud as HBs particles into the lumen of the vesicles. Both the large and small surface proteins (but not MHBs) are required for virion formation (Bruss and Ganem, 1991). After budding of particles, cytosolic domains are enclosed within the particles, while the ER luminal domains are exposed to the surface. In agreement with this topological requirement, the preS2 domain of MHBs and the ER luminal portion of HBsAg are at the surface of HBsAg particles. The preS1 domain is partially translocated from the interior side of the viral envelope to the surface during secretion (Bruss *et al.*, 1994). Infected cells secrete, in addition to virions, large amounts of subviral noninfectious 22 nm particles. If the proportion of LHBs is low, 22 nm spheres are secreted; if it is moderate, filaments are secreted; and if it is higher than that of HBsAg and MHBs, the particles are retained in the ER (Standring *et al.*, 1986).

1.2.2 Outcomes of HBV infection

Infection with HBV is associated with a wide spectrum of liver damage ranging from fulminant acute hepatitis to chronic HBV with minimal hepatic change or liver cirrhosis and/or hepatocellular carcinoma, HCC. Patients with the different forms of chronic infection are variously termed asymptomatic carriers (ASC), chronic persistent (or simple) carriers (CPH) and chronic active (or super) carriers (CAH). Although hepatocytes are the only site of productive infection, replicative forms of HBV have been found in the following immunologically privileged sites: lymph nodes, spleen, kidney, pancreas, brain and in some endocrine tissues, including testis, ovary, adrenal, and thyroid glands (Yoffe *et al.*, 1990).

1.2.2.1 Acute self limiting hepatitis B

The host immune response critically determines the clinical presentation and outcome of HBV infection. A vigorous HLA class II-restricted CD4+, T helper cell (Th) response to multiple epitopes in the nucleocapsid antigens (HBcAg and HBeAg) is detectable in the peripheral blood of almost all patients with self-limited acute hepatitis although their class II-restricted, envelope-specific response is much less vigorous (Ferrari *et al.*, 1990). Table 2 lists published core, polymerase, surface and X antigen linear epitopes recognized by T-cells. The residue core-50-69 is most commonly recognized by acutely infected patients

Table 2. Published class I- and class II-restricted epitopes of HBV

Epitope	genome aa	Published Epitope	HLA	Class	Source	Author
preS2-44-53	41-50	SIESRTGDPAP	A2	I	AVH PBL	Chisari and Ferrari, 1995
HBs-14-22	66-74	VLQAGFFLL	A2	I	AVH PBL	Chisari and Ferrari, 1995
HBs-20-28	72-80	FLLTRILTIP	A2	I	AVH PBL	Chisari and Ferrari, 1995
HBs-20-28	72-80	FLLTRILTI	DPw4	II	VAC PBL	Celis *et al*, 1988
HBs-20-32	72-80	FLLTRILTIPQSLD	DR2w15	II	VAC PBL/ CAH liver	Bainaba *et al* 1994
HBs-41-49	93-101	FLGGTPVCL	A2	I	AVH PBL	Chisari and Ferrari, 1995
HBs-89-96	140-148	LLCLIFLLV	A2	I	AVH PBL	Nayersina *et al*, 1993
HBs-97-106	149-158	LLDYQGMLPV	A2	I	AVH PBL	Nayersina *et al* 1993
Pol455-463	150-158	GLSRYVARL	A2	I	AVH PBL	Reherrmann *et al* 1995a
HBs-158-172	210-224	FADPIWEWASVRFS	ND	I	ASCPBL	Shi *et al*, 1994
HBs-172-180	224-232	WLSLLVPFV	A2	I	AVH PBL	Nayersina *et al*, 1993
HBs-185-194	237-246	GLSPTVWLSV	A2	I	AVH PBL	Nayersina *et al*, 1993
HBs-207-216	259-268	SIVSPFIPLL	A2	I	AVH PBL	Nayersina *et al*, 1993
Pol 575-583	270-278	FLLSLGIHL	A2	I	AVH PBL	Reherrmann *et al* 1995a
Pol773-782	468-477	ILRGTSFVYV	A2	I	AVH PBL	Reherrmann *et al* 1995a
X-44-53	502-511	AVFADHGAHL	A2	I	CPH PBL	Chung *et al*, 1999
X-42-60	510-128	HLSLRGLPV	A2	I	CPH PBL	Chung *et al*, 1999
Pol-816-824	511-519	SLYAVSPSVFSHL	A2	I	AVH PBL	Reherrmann *et al*, 1995a
X-91-100	549-558	KVLHKRTLGL	A2	I	CPH PBL	Chung *et al*, 1999
X-92-100	550-558	VLHKRTLGL	A2	I	CPH PBL	Chung *et al*, 1999
X-99-108	557-566	GLSAMSTTDL	A2	I	CPH PBL	Chung *et al*, 1999
X-115-123	573-581	CLFKDWEEL	A2	I	CPH PBL	Chung *et al*, 1999
X-126-134	584-592	EIRLKVFVL	A2	I	CPH PBL	Chung *et al*, 1999
X-133-141	591-599	VLGGCRHKL	A2	I	CPH PBL	Chung *et al*, 1999
CORE-1-20	635-654	MDIDPYKEFGASVELLSFLP	ND	II	AVH PBL	Ferrari *et al*, 1991
CORE-18-27	652-661	FLPSDFFPSV	A2	I	AVH PBL	Bendelli *et al*, 1993
CORE-38-47	662-681	RDLLDTASALYREALESPEH	DR7/DP0402	II	AVH PBL	Ferrari *et al* 1991
CORE-60-69	684-703	PHHYALRQAILCWGELMTLA	ND	II	AVH PBL	Ferrari *et al* 1991
CORE-87-100	721-734	SYVNTNMGLKIRQL	A11	I	HCC	Tsai *et al*, 1996
CORE-111-125	745-759	GRETVLEYLVSFGVW	DR1	II	AVH PBL	Chisari and Ferrari, 1995
CORE-117-131	751-765	EYLVSFGVWIRTPPA	DRw52/DR6	II	AVH PBL	Ferrari *et al*, 1991
CORE-120-139	754-773	VSFGVWIRTPPAYRPPNAPI	DR	II	AVH PBL	Ferrari *et al*, 1991
CORE-141-151	775-785	STLPETTVVRR	A31/Aw68	I	AVH PBL	Missale *et al* 1993
Pol-61-69	832-840	GLYSSTVPV	A2	I	AVH PBL	Reherrmann *et al* 1995a
preS1-21-28	972-979	VPNPLGFFPDH	A11	I	VAC PBL	Jin *et al*, 1988
preS1-24-33	972-979	PLGFFPDHQL	DQ5	II	VAC PBL	Ferrari *et al*, 1992
preS1-29-48	980-999	QLDPAFRANSNNPDWDFNPN	DR4/DR5	II	VAC PBL	Ferrari *et al*, 1992
preS1-81-95	1032-1047	AQGLTTVPAAPPA	DR4/DR5	II	VAC PBL	Ferrari *et al*, 1992
preS1-94-108	1045-1059	PASTNRQSGRQPTPI	DR5	II	VAC PBL	Ferrari *et al*, 1992
preS2-1-15	1071-12	MQWNSTTFHQALCDP	A3	I	CAH LIVER	Bainaba *et al*, 1989
preS2-1-15	1071-12	MQWNSTTFHQALCDP	DR2	II	CAH LIVER	Bainaba *et al*, 1989

irrespective of their HLA backgrounds (Ferrari *et al.*, 1990; 1991).

The HBV-specific humoral response is not an important mediator of liver cell injury nor is it sufficient to resolve an acute or chronic infection. Antigen levels are, however, an important indicator of underlying T-cell function, or dysfunction, and predict the degree of T-cell sensitization to the antigen being assayed and the predominance of a particular Th-cell phenotype (Th1 versus Th2). In the murine system, certain HBeAg-specific T-cell site/major histocompatibility complex (MHC) combinations preferentially elicit either a Th1-like (proliferation) or a Th2-like (antibody) response. The Th2 subset which favours higher antibody levels predominates in chronic disease over cell-mediated immunity and a Th1 subset in acute disease (Milich *et al.*, 1995). The profile of serological markers detected at any one time reflects the time of onset of infection and this is judged against the typical patterns shown in Figure 10a. Screening of HBsAg, IgG antibody to HBcAg (anti-HBc) and antibody to HBsAg (anti-HBs) by standard radioimmunoassay (RIA) or enzyme immunoassay (EIA) is usually sufficient to diagnose acute HBV infection and this initial screen is only extended to include HBeAg, IgM antibody to HBcAg (IgM anti-HBc) and antibody to HBeAg (anti-HBe) when further disease evaluation is necessary (Hollinger, 1990; Sim, 1993).

1.2.2.2 Chronic HBV infection

Five to ten percent of acutely infected adults and 90 percent of neonatally infected infants become persistently infected by HBV and develop chronic liver disease of varying severity. Neonatal tolerance is important in infection during gestation (Milich *et al.*, 1990) but the basis for tolerance during adult onset infection is less understood. Various studies have shown that the final outcome of infection is dependent on the specific viral isolate, the sex of the host, age at time of infection, infectious dose of the virus, human leucocyte antigen (HLA) background of the host and subsequent host immune response. Ultimately the dominant cause of viral persistence is a weak antiviral immune response to the HBV antigens, but the determinants that control viral virulence or host susceptibility (or resistance) in the natural host are not yet understood fully in terms of the molecules or genes involved. Selection of mutants with variant T-cell epitopes is low and unlikely to be responsible for HBV persistence since the cytotoxic T lymphocyte (CTL) response is against multiple epitopes in acute hepatitis B (Table 2; Rehermann *et al.*, 1995b). On the

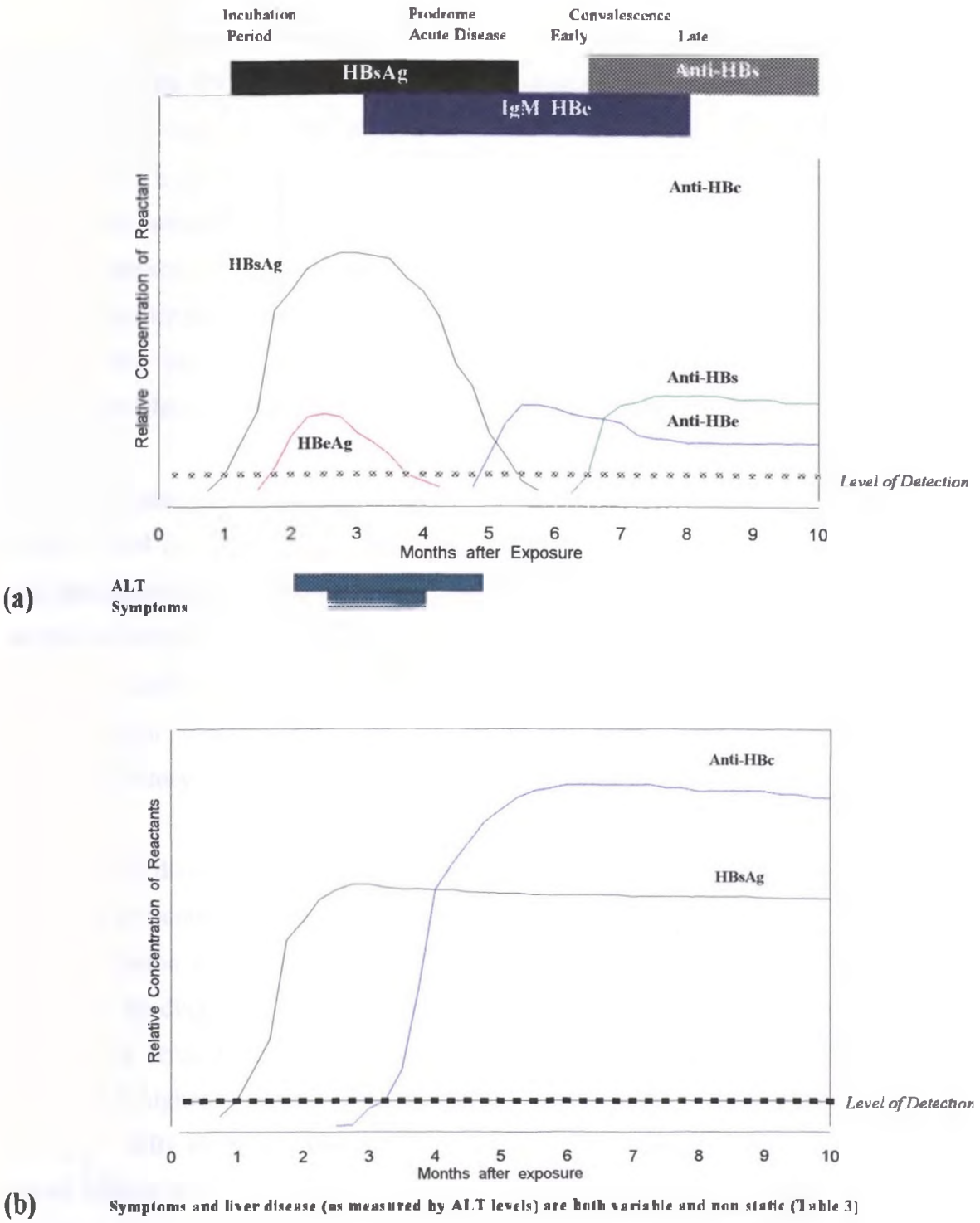


Figure 10. Serological and clinical patterns associated with HBV infection (a) with resolution [adapted from Hollinger, 1990] and (b) without resolution [adapted from Sim, 1993]

ALT: alanine aminotransferase; HBeAg: hepatitis B virus e antigen; HBsAg: hepatitis B virus S antigen; anti-HBc: IgG antibody to hepatitis B core antigen; anti-HBe: antibody to hepatitis B e antigen; anti-HBs: antibody to hepatitis B S antigen; IgM anti-HBc: IgM antibody to hepatitis B core antigen

other hand, in the context of a preexisting ineffective T-cell response to HBV, or if the CTL repertoire to HBV is narrow, the emergence and dominance of such mutants can inactivate viral epitopes or convert them into active T-cell antagonists. This occurs when substitutions of T-cell receptor (TCR) contact sites create analogue peptides that can still interact with the T-cell receptor but are not able to deliver a full stimulatory signal. Natural variants of the core-18-27 CTL epitope that interfere with recognition of the wild type epitope and act as T-cell receptor antagonists have been identified in chronic carriers (Bertoletti *et al.*, 1994). If individual infected cells simultaneously express both variant and wild type peptides, the molecular requirements for antagonism would occur. This effectively protects the cell containing the mutant from immune elimination by CTL specific for the wild type epitope and leads to selective expansion of the variant viral population.

Lymphocyte recognition can also be evaded by downregulation of the expression of viral proteins. Viral integration interrupts HBV coding regions and downregulates some of the viral proteins which reduces their visibility to the immune system. Conversely, the core antigen can suppress the β-interferon gene (Twu *et al.*, 1988; Whitten *et al.*, 1991) and polymerase inhibits cellular response to α- and γ-interferon (Foster *et al.*, 1991), thus downregulation would also affect the ability of HBV to downregulate the HLA immunoregulatory molecules and the accessory molecules which they induce.

As with acute disease, definitive diagnosis of chronic carriage of HBV is dependent upon serological screening and the interpretation of the possible combinations of markers in the serum in relation to the clinical stage of infection (Figure 10b). Although the majority of patients can be diagnosed using the standard serological screens there is often a delay in the detection of seroconversion. Sera from patients with chronic type B hepatitis show significantly higher levels of free anti-HBe, HBeAg/anti-HBe immune complexes and HBsAg/anti-HBs immune complexes compared to acute hepatitis patient sera. For these reasons Milich *et al.* (1995) used a direct binding EIA to detect anti-HBe and anti-HBs antibodies in the presence of antigens in the serum, and used these to define three serological profiles (Table 3) associated with the three different forms of the chronic carrier state.

The different forms of chronic carriage are not static and one form of chronic disease may

Table 3. The three serological profiles of HBV chronic carriage*

Clinical diagnosis	Frequency	Serological profile	HBeAg-Th	Functions	Liver disease
ASC	44.4%	I. ▪ Decreasing anti-HBc IgG1 ▪ HBV DNA ++	none		none
	55.6%	II. ▪ Increasing (↑) anti-HBc IgG1 ▪ ↑ anti-HBe IgG1 ▪ ↑ anti-HBenv ▪ HBV DNA ++	Th2-like	increased antibody production	none
CAH CPH	±100%	III. ▪ rapidly increasing anti-HBc IgG1 and IgG3 ▪ ↑ anti HBe IgG1 and IgG3 ▪ ↑ anti-HBenv ▪ HBV DNA + to -	Th1-like + Th2-like	increased inflammatory response anti-HBc/ HBe switched to G3	++ to ±

* adapted from Milich *et al*, 1995

Chronic carriage has been defined as asymptomatic chronic carriage (ASC), chronic active hepatitis (CAH) and chronic persistent hepatitis (CPH) based on the whether disease was characterized by increased inflammation (T-helper cell 1 like [Th1-like]) or increased antibody production (T-helper cell 2 like [Th2-like]). The trend and type of the antibody response and the levels of HBV DNA circulating in the serum further enable one to classify the severity ([++] severe to [±] moderate liver disease) and stage of liver disease.

progress to a more aggressive form and vice versa. Chronic carriers may have periodic episodes of reactivation of disease during which accentuated nucleocapsid-specific T-cell responses are preceded by increased serum HBV-DNA and HBeAg which subsequently drop dramatically as the flare activity subsides. During reactivation serum HBsAg concentrations remain unchanged and the proliferative response to HBV envelope antigens remains weak or undetectable (Maruyama *et al.*, 1993). There is also evidence for the existence of a serologically silent persistent carrier state of hepadnavirus in apparently healthy individuals with a past history of a self-limited episode of acute hepatitis (Michalak *et al.*, 1994).

1.2.2.3 Hepatocellular carcinoma, HCC

The epidemiological evidence implicating HBV in the genesis of HCC is overwhelming (Beasley, 1981) but the precise role which it plays is still unclear. HBV encodes three known *trans*-activating products. The X gene product and the MHBs possess transcriptional activator functions, while the C gene has been shown to be a suppressor. X also has many

other functions which taken together have the potential to favour selection of transformed cells (discussed in paragraph 1.1.3.4). On top of this the constant hepatocellular damage caused by chronic disease can result in hyperplasia which even further raises the risk for increased mutation rates.

In the majority of HBV-associated HCC, the viral DNA is integrated into the cellular chromosomal DNA. Integration is a frequent event which must occur soon after infection since it has been observed in acutely infected cells. Integration into the chromosome is random, although chromosomes 11 and 17 seem to be represented at a higher frequency (Tokino and Matsubara, 1991). In some cases the site of integration may predetermine cell transformation, 20 percent of the tumours involving WHV DNA have viral DNA inserted within one of the two N-*myc* genes (Fourel *et al.*, 1990), although no known oncogene has been associated with integrated HBV DNA. The favoured site of HBV integration is via the direct repeat sequences, DR1 and DR2, and the disruption of this region leads to important sequelae since in most tumours the X gene is truncated and linked to cellular DNA. This often qualitatively alters the activational capabilities of X (Wei *et al.*, 1995). In the same way the coding region of cellular genes could be dissociated from their own regulatory elements and brought under the influence of viral enhancers and promoters leading to *cis*-activation of cellular genes by viral genes (and vice versa) or, alternatively, greater stability of the transcripts.

HBV also mediates genomic imbalances and rearrangement of both viral and cellular DNA including inversion, duplication, deletion and chromosomal translocation (Pineau *et al.*, 1996; 1998) and is reported to enhance DNA recombination (Hino et al., 1991). Clearly the diversity and multiplicity of the molecular effects of integrated HBV DNA on the hepatocyte contribute substantially to multistep hepatocarcinogenesis.

Other extraneous factors are also thought to cooperate with HBV in tumourigenesis in ways which are not always well defined. Although these factors are important, in the absence of HBV there is little evidence to show that they significantly influence hepatocarcinogenesis in high HBV-endemic regions. Dietary contaminants such as aflatoxin (Bressac *et al.*, 1991), hormones and agents which contribute to hepatic damage have all been implicated in the pathogenesis of HCC. In both humans and experimental mice systems, the incidence of

HCC is significantly higher (or more rapid) in males. A glucocorticoid responsive element present on the viral DNA may partly be the basis of this hormonal bias (Tur-Kaspa *et al.*, 1988).

1.2.3 Common mutants of HBV and their effect on pathogenesis

Viral evolution can be influenced by antiviral and other therapies or the immune response in a particular individual. Although mutants may play a role in establishing chronicity or fulminant hepatic failure, only a few examples exist that show a mutant HBV strain which has a direct impact on the course of the liver disease. Mutants can play a role in pathogenesis by altering protein function or the rate of transcription and/or replication. They may change host range or cell tropism or be responsible for vaccine escape. The following are some of the most commonly reported variants of HBV.

1.2.3.1 Precore mutants

A single base substitution resulting in a new translational stop codon within the precore open reading frame blocks the synthesis of the HBeAg. Since common precore mutations stabilize the stem-loop structure of the pgRNA, which serves as the origin of viral replication and as the packaging signal, mutants could gain a replicative advantage which would lead to their eventual dominance over other viral quasi-species present in the infected patient.

1.2.3.2 Core promoter mutants

The double mutation of nucleotides 1762 and 1764 in the core promoter/enhancer II region (A to T and G to A, respectively) is frequently observed in HBV sequences isolated from chronic patients and immunosuppressed patients with severe liver disease (Takahashi *et al.*, 1995). The mutated core promoter can no longer bind a liver-enriched transcription factor and this reduces the transcription of precore RNA and reduces the expression of e antigen. The reduction of precore gene expression is accompanied by an increase in progeny virus production. This increase occurs at or just prior to encapsidation of the pgRNA and explains why HBV containing this double mutation becomes the predominant quasi-species during chronic infection.

Baumert *et al.* (1996), while investigating viral strains from five patients with fulminant

hepatitis, found a mutation in the core promoter region which leads to enhanced encapsidation and upregulation of HBV replication. This double mutation involved a C to T at nucleotide position 1768 and a T to A at nucleotide position 1770. They showed that pgRNA transcription had not been significantly altered but that the mutations had a *trans* effect on the encapsidation of pgRNA. Since the mutations had no effect on the *trans*-activation activities of HBx, the authors suggest that a novel genetic element may be involved in HBV encapsidation.

A different variant with mutations in the core promoter region resulting in fulminant hepatitis was reported by Pult *et al.* (1997). The genome of this variant was 3,100 bp due to two in-frame deletions of 18 bp and 108 bp, respectively, but also had an 11 bp insertion which created a putative HNF1 binding site in the core promoter region and resulted in a truncated X protein. Transfection experiments indicated a higher replication competence compared to wild type HBV DNA and removal of the HNF1 binding site reversed the level of replication to that of wild type HBV. Gel retardation assays using recombinant HNF1 identified the insert as a *bona fide* HNF1 binding site.

1.2.3.3 PreS1 deletion mutants

PreS1 deletion mutants occur in 4 to 10 percent of chronic hepatitis B infected individuals. These mutants have been found to replicate at a substantially higher level than wild type HBV (Melegari *et al.*, 1997) but do not produce the 2.1 kb mRNA and thus lack SHBs. *Trans*-complementation with constructs expressing all three S proteins restored secretion by providing an envelope for the mutant virion. PreS1 mutants augment the pool of cccDNA providing additional template for synthesis of viral protein which renders hepatocytes a better target for cytotoxic T-cells and promotes chronic liver injury.

1.2.3.4 Core internal deletion mutants

Internal core deletion mutants are frequently found (Yuan *et al.*, 1998). The size of the deletions vary from 18-61 amino acids and are most often located within amino acid 80 to 120 of the core antigen. These variants have been found in asymptomatic carriers, chronic active hepatitis and hepatoma patients. The deleted domain correlates with immunodominant T-cell epitopes suggesting an immune escape nature of this deletion. They are replication and packaging defective but can be rescued by wild type HBV via *trans*-

complementation. The deleted core was highly unstable *in vivo*, though it appears to polymerize readily *in vitro*.

1.2.3.5 Variants containing spliced genomes

Analysis of the large numbers of smaller genomes obtained from full genome amplification of HBV revealed genomes with various deletions as well as genomes derived from reverse transcription of spliced mRNAs (Günther *et al.*, 1997).

1.2.3.6 Surface antigen escape mutants

Mutations in the "a" epitope of the surface antigen of HBV can lead to patients with chronic liver disease who are positive for antibodies against the core antigen but have no detectable HBsAg. The most common mutation is at amino acid 145 and is postulated to evolve during vaccination (Carman *et al.*, 1990). Changes to the cysteines which are necessary to form the SHBs loops also affect the three dimensional properties of HBsAg.

The emergence of "escape" mutants under selective pressure of a polyclonal anti-HBs in transplant recipients is also a problem as they can reinfect an individual in spite of anti-HBs administration against predominant strains.

1.2.4 Model of HBV Infection

The model summarizes the immunopathogenesis of HBV (Figure 11; Chisari and Ferrari, 1995). Antiviral CD8+ CTL play a major role in eradication of HBV infection by virtue of their capacity to identify and kill virus-infected cells through recognition of endogenously

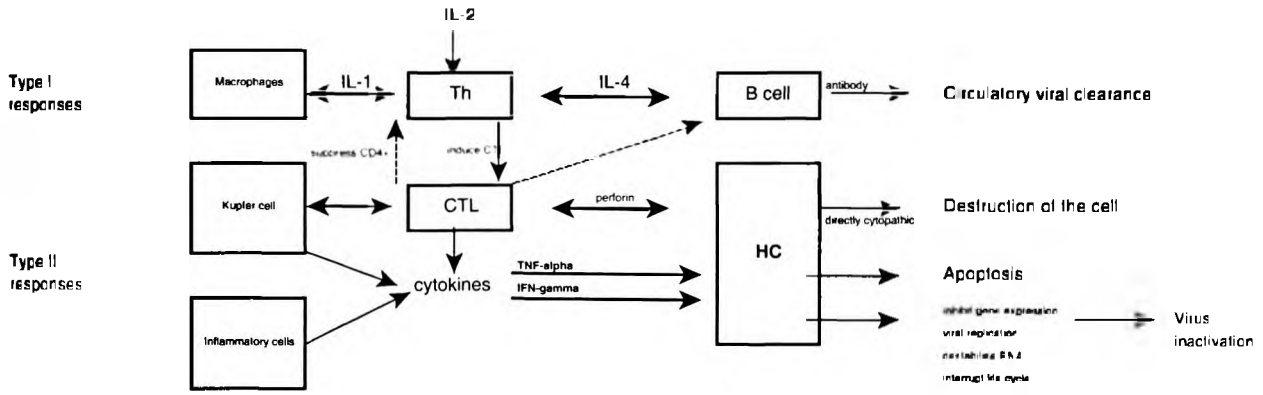


Figure 11. Model of HBV immunopathogenesis (adapted from Chisari and Ferrari [1995])
Anti-viral CD8+ CTL identify and kill virus-infected cells through recognition of endogenously synthesized HBV antigen presented to HLA class I molecules at the hepatocyte (HC) membrane. They also inhibit replication by secretion of the cytokines IFN-γ and TNFα triggering their target HC to undergo apoptosis while HBcAg-specific CTL suppress the CD4+ proliferative T-cell response. The class II-restricted nucleocapsid-specific Th response induces virus-specific CTL and helps HBV envelope-specific B cells to produce neutralizing antibodies.

synthesized HBV antigen presented to HLA class I molecules at the hepatocyte (HC) membrane. The CTL not only destroy the hepatocyte directly but also inhibit HBV replication non-cytopathically by secreting interferon gamma (IFN-γ) and TNFα following antigen recognition. Cytokines non-cytolytically inhibit HBV gene expression and viral replication, destabilize RNA, interrupt viral life cycle and result in viral clearance without destroying infected liver. Suppression of HBV gene expression and replication by the CTL response can work against the host and contribute to viral persistence by allowing infected cells to escape immune recognition. HBsAg-specific CTL can also trigger their target hepatocytes to undergo programmed cell death (apoptosis) or selectively kill a variety of cell types, including HBsAg-specific B cells following antigen uptake via the immunoglobulin receptor, owing to the remarkable ability of exogenous soluble HBsAg to enter the class I-processing pathway. Intrahepatic HBcAg-specific CD8+ T-cells,however, specifically suppress the CD4+ proliferative T-cell response to HBV in patients with chronic hepatitis B (Ferrari *et al.*, 1987). Anergic Th1 clones specific for HBV core peptides also inhibit other HBV core-specific CD4+ T-cells *in vitro* (Diepolder *et al.*, 1996). Furthermore CTLs recognize HBV surface antigen on the surface of Kupffer cells in the liver, following which they secrete Th1 cytokines that mediate their pathogenic and antiviral effector functions at the level of the hepatocyte.

In addition non-specific inflammatory cells amplify cytopathic effects and worsen severity of disease via the cytokines they secrete.

The class II-restricted nucleocapsid-specific Th response also plays a critical role in viral clearance contributing in several ways. It induces virus-specific CTL and it helps HBV envelope-specific B cells to produce neutralizing antibodies. Th cells may also play an important effector role in HBV pathogenesis and viral clearance since disease severity correlates with the number of activated T-cells in the liver. Although endogenously synthesized viral antigen does not associate efficiently with HLA class II molecules inside the cell, HBV enters the class II-processing pathway in T-cells via uptake by the transferrin receptor (Franco *et al.*, 1992). Intrahepatic macrophages also express high levels of HLA class II and release their own cytokines.

A positive treatment response correlates with a significant increase in the levels of TNFα

and IL-1β, both potent antiviral substances produced by the patient's monocytes. Since core protein suppresses TNFα, it could contribute to the development of HBV chronic persistence by inhibiting cellular antiviral responses, and thus also increase the risk of hepatocyte transformation.

CHAPTER 2

INTRODUCTION: RATIONALE FOR THE STUDY

2.1. Motivation for the study

Infection with HBV can result in a variety of diseases ranging from minor to severe acute hepatitis. In 10 percent of adult patients the virus persists leading to chronic interferon-sensitive or interferon-resistant hepatitis. Chronic carriage also presents a wide spectrum of sequelae from asymptomatic to aggressive and can even alternate between quiescence and more active liver disease. Not surprisingly therefore, HBV persistence also causes varying degrees of liver damage and cirrhosis which can ultimately result in HCC. The range of response to infection can only be explained if different viruses exist which differ in their virulence, or if host responses vary, or both. Recent advances in molecular biology and immunology reveal a varying and complex interaction between virus and host which cannot be explained by viral mutation or host genetics alone suggesting an active relationship between virus and host (Foster *et al.,* 1993).

Safe, immunogenic and protective plasma-derived hepatitis B vaccine has been available in the USA since 1982 (Szmuness *et al.,* 1980) and a recombinant vaccine of equal safety and efficacy has been registered since 1987 (Scolnick *et al.,* 1984) and the year 1995 saw the start of mass immunization of infants against HBV in South Africa. However, even at the level of vaccine response the HBV/host interaction is complex and 5 to 14 percent of immunocompetent adults immunized with hepatitis B vaccine do not acquire anti-HBs (Katkov *et al.,* 1994). Moreover, and particularly in the light of the severity of the HIV epidemic in sub-Saharan Africa, the large majority of immunocompromised persons do not develop antibody after vaccination.

There are an estimated 385 million chronic carriers of HBV worldwide and a significant number (75 million; Hilleman, 1994) of these people will die either from liver failure or HCC. In sub-Saharan Africa the overall HBsAg carrier rate is 5 to 20 percent, ranking as

one of the highest carrier rates in the world. Among Mozambiquan Shangaans the age-standardised incidence rate of HCC is 103.8 per 100,000 per year (Kiire, 1990). The approved treatment for chronic type B hepatitis is alpha-interferon therapy but, apart from the expense of such treatment, only 30 to 40 percent of patients respond and relapse is reportedly common when treatment is stopped (Terrault, 1996). At present there are no pretreatment characteristics that predict whether or not a given patient will respond or relapse upon cessation of treatment (Gerin, 1992). Thus despite developments in treatment and prophylaxis HBV remains one of the most important worldwide public health problems with special relevance in the southern African context.

Very little is known of the serotypes or genotypes of HBV circulating in southern Africa. A study of the world distribution of HBsAg subtypes performed in 1983 using counter-electrophoresis and immuno-diffusion techniques included 200 HBV positive specimens obtained from the Kwazulu-Natal province of South Africa and reported a 97.7 percent frequency of serotype *adw* together with 2.3 percent serotype *ayw* in the eastern region of South Africa (Courouce-Pauty *et al.*, 1983). Prior to publication of my initial results (Bowyer *et al.*, 1997), the only southern African sequence data available came from the partial sequencing of a Zimbabwean isolate (Chirara *et al.*, 1994) and the sequence of integrated HBV and human flanking regions in the genome of the PLC/PRF/5 cell line derived from liver tissue of a Mozambiquan carrier with HCC (Rivkina *et al.*, 1988).

A complete understanding of how the HBV virus evolves must of necessity include studies of HBV sequences from all geographic regions and, in particular, the regions, such as sub-Saharan Africa where the virus is endemic. The main aim of this study was to identify the strains of HBV circulating in southern Africa and to characterize the genotypic partitions to which they belong. As data from regions such as southern Africa gradually begin to complete the global picture, it is hoped that the enigmas of HBV persistence and resistance to therapy and vaccination will be explained, enabling the design of more effective vaccines and better strategies to combat HBV disease. At the same time the background data provided by this study will serve as a prelude to future research providing the baseline against which the success of treatment and prevention programmes can be monitored and measured.

2.2. Aims and organization of the study

AIM: To identify the strains of hepatitis B circulating in southern Africa and to determine whether variations exist within the infecting virus and infected populations in this part of the world.

To achieve these goals the project was divided into 7 separate sections which are outlined briefly below. Each study concentrated on one aspect of the southern African picture.

1. The prevalence and molecular epidemiology of HBV infection in southern Africa. This study reviewed local serological data including a recent random cross sectional study of first time blood donors from all provinces of Namibia carried out in conjunction with the Namibian Blood Transfusion Service (NBTS). HBV genotypes circulating in southern Africa were then determined using PCR fragment size and RFLP analyses adapted to differentiate local strains. Genotypic data was confirmed and extended by sequence analysis as well as genotype-specific EIA assay.

2. The preS2 and HBsAg genes of representative South African specimens identified in the screening assays were sequenced and compared with global sequence data using phylogenetic analysis to determine the degree of uniqueness of locally circulating strains.

3. In order to map variant nucleotides along the HBV genome a database of 65 full genome HBV sequences representing each of the 6 genotypes of HBV was created, manipulated and analysed. The effect of the nucleotide variation on the HBV proteins was also examined in all three reading frames. Results were used to define differences characteristic of the genotype partitions and to separate this change from more random mutation, or non-definitive change, between individual specimens. A site was defined as a "hotspot" if it varied in greater than five percent of the database specimens.

4. Phylogenetic analysis had revealed several specimens which switched genotype when different regions of the genome were examined suggesting the existence of HBV recombinants. In order to explore this phenomenon I compared the relationships within and between the genotypes of HBV, 60 nucleotides at a time, at regular concomitant intervals along the genome and also compared each of 65 full genomes with the consensus sequence

sequence of each of the genotypes of HBV. Anomalous regions of the genomes were mapped and characterized.

5. The HLA-DR type of a small cohort of HBsAg-positive patients from South Africa and Namibia were compared with published population frequencies for ethnic groups in southern Africa and HLA types associated with HBV disease progression.

6. The class II core immune response of the HLA-typed southern African patients was localized by proliferation assay using overlapping core peptides and examined in the context of patient DR type, published linear epitopes and predicted HLA-DR binding motifs within immunogenic regions.

7. The core gene of selected local specimens was compared with GenBank core gene sequence data with particular reference to the similarities and differences at published linear epitopes and predicted HLA-DR binding motifs.

CHAPTER 3

MATERIALS AND METHODS

3.1. Patients

Specimens were obtained from both rural and urban populations of South Africa and selected rural populations of northern Namibia. Global comparisons were done using HBV sequence data from GenBank. The full list of GenBank accession numbers accessed, as well as local specimens for which sequence data (published or unpublished) is available is given in Appendix B, Table 25. This table includes information regarding the size and position of the region of each specimen sequenced. Unsequenced specimen data is listed in the Materials and Methods sections of relevant chapters only.

3.1.1 South African Patients

3.1.1.1 Urban cohort

The urban South African sample consisted of 37 HBsAg-positive specimens selected from the thousands of specimens sent annually to the National Institute for Virology (NIV) for routine diagnostic HBV serology, and 19 patients with acute HBV diagnosed in hospitals in the Gauteng area. The preS2 region and/or HBsAg region of 28 and 27 specimens respectively was sequenced and the data submitted to GenBank (Accession numbers U87725-U87748, U87851-U87852, U88094-U88096). Unpublished complete (ES01-ES05) or partial (E08, E17, O01, O10, P08 and S09) HBV core sequence data is available for 11 of these specimens.

3.1.1.2 Rural cohort

The rural South African cohort consisted of the 238 HBsAg-positive community infected children identified within a large randomly selected cohort from the Eastern Cape Province. The children were between the ages of 0 and 72 months and the specimens were collected to determine the baseline prevalence of HBV infection in this region prior to the introduction of mass immunization of infants against HBV in South Africa (Vardas *et al.*, 1999). A

random sample of 57 of these HBsAg-positive specimens were selected for genotypic screening. However, since only finger or heel prick blood specimens were taken from the children, only 37 (64.9 percent) of the selected group yielded sufficient and/or suitable material for the digestions.

3.1.2 Namibian Patients

The Namibian specimens were obtained from three separate cohorts. Two of these were provided by the Namibian blood transfusion service (NBTS) while the third was a small sampling of HBsAg- and HBeAg-positive specimens from Ovamboland.

The first blood donor specimens supplied to the NIV by the NBTS consisted of a pilot study of 230 specimens collected from the North-West (Ovamboland, Ondangwe and Oshakati) and north east (Rundu region) regions of the country. Thirty-nine of these specimens were found to HBsAg-positive and 16 of these had sufficient volume to allow genotype determination.

A selection of 32 HBsAg- and HBeAg-positive specimens from the Ovambo region of north west Namibia, collected in 1984 for an HBV transmission study (JB numbers, Botha *et al.*, 1984), formed the second rural sample. Five of the 32 were mother/child sets. Twenty-four of the specimens were successfully typed. The complete core gene of 9 of these specimens (JB01, JB02, JB04, JB05, JB07, JB11, JB13, JB16 and JB17) was sequenced.

Lastly whole blood from a group of 8 HBsAg-positive voluntary blood donors (N231-238) from the North East who had been diagnosed by NBTS were genotyped and included in the proliferation assay. Seven of the 8 specimens were also HLA-DR typed. The eighth specimen, N233, was repeatedly negative in the HLA-DR polymerase chain reaction (PCR) and was therefore excluded from the later studies.

3.1.3 GenBank specimens

Apart from the local specimens described in paragraphs 3.1.1 and 3.1.2, sequence data from a total of 179 GenBank specimens (some representing horizontal specimens from the same patient) and 22 sequence data listings in the literature were used, together with local sequence data, in phylogenetic analyses and global comparisons (Appendix B).

3.2 Molecular Methods

3.2.1 Viral DNA extraction

One hundred and twenty-five microlitres (µl) of serum was mixed with 375 µl lysing solution (2 percent SDS, 50 millimolar (mM) EDTA, 200 mM NaCl and 100 mM Tris-HCl pH 8.2) containing 10 µg/µl of Proteinase K and incubated at 65 °C for one hour (or 37 °C overnight). Nucleic acids were extracted twice with an equal volume (500 µl) of Tris-buffered phenol (pH 8.0)/chloroform (1:1) and then precipitated with 1/10 volume of three Molar (M) sodium acetate (pH 6.0) and 2½ volumes of 100 percent ethanol. The tube was placed at -70 °C for 30 minutes and then centrifuged at 10,000g for five minutes. The supernatant was discarded and the pellet washed twice in 70 percent ethanol and then dried under vacuum for 10 minutes. The pellet was resuspended in 20 µl of sterile water.

Alternatively, a commercial DNA purification kit (QIAamp blood kit, QIAGEN GmbH, Germany) was used to extract the DNA directly from 200 µl of serum, plasma or peripheral blood mononuclear cells (PBMCs) according to the manufacturer's intructions. DNA was eluted in 100 µl of buffer AE provided with the kit.

3.2.2 Polymerase Chain Reaction (PCR)

A robust, in-house PCR was developed by optimising and refining published HBV PCR methodology (Shih et al., 1991). Final reagent concentrations were: 1X Roche Diagnostics (Mannheim, Germany) PCR Reaction buffer, 1.5 mM $MgCl_2$, 200 µM of each deoxynucleoside-triphosphate (dNTP), 1 micromolar (µM) of both forward or sense (Sn) and reverse or anti-sense (Asn) primers (see Table 4 for primers used and amplicon sizes expected from various primer pair combinations) and 2.5 units of *Taq* Polymerase (Roche Diagnostics, Mannheim, Germany). Reagents and reaction mixtures were kept on ice throughout to prevent non-specific priming and subsequent non-specific amplification artifacts. Ten microlitres of extracted HBV DNA in a final reaction volume of 100 µl was subject to the following cycling conditions in a Biometra trioblock. Denaturation at 94 °C for two minutes followed by 36 cycles of denaturation at 94 °C (one minute), annealing at 48 °C (one minute) and amplification at 72 °C (two minutes). A final elongation step (three minutes at 72 °C) ensured that all PCR products were full-length and complete. Amplified

Table 4. (a) HBV-specific primer sequences and (b) resultant amplicon sizes obtained using appropriate forward (Sn: sense) and reverse (Asn: antisense) primer pairs

(a)

Primer	Primer sequence (5' to 3')
Sn1738	AGT TGG GGG AGG AGA TTA G
Sn1821 (P3)	[CTC GCT CGC CCA AA]T TTT TCA CCT CTG CCT AAT CA
Sn2833	CTT GGG AAC AAG AGC TAC AGC AT
Sn3088	CCT CAG GCT CAG GGC ATA
Asn256	AGA AAA TTG AGA GAA GTC CAC CAC G
Asn688	AGC CCT ACG AAC CAC TGA ACA AAT
Asn1018	CCA CAT TGT GTA AAT GGA GCA GC
Asn1806 (P4)	[CTG GTT CGG CCC] AAA AAG TTG CAT GGT GCT GG
Asn2467	GTA AAG TTT CCG ACC TTA TGA GTC C
Asn2973	GGG TTG AAG TCC CAA TCT GGA

(b)

Sn\Asn	Asn256 Tm 61.3	Asn688 Tm 61.01	Asn1018 Tm 60.64	Asn1806 (P4) Tm 55.25	Asn2467 Tm 61.3	Asn2973 Tm 58.28
Sn1738 Tm 56.66			**Full core + surface** (2485-2524)		**Full core** (748-754)	
Sn1821 (P3) Tm 53.97				**Full genome** (3182-3221)		
Sn2833 Tm 60.72	**PreS** (635-668)	**Early surface** (1067-1100)	**Full surface** (1396-1429)			**PreS1** (128-161)
Sn3088 Tm 58.24	**PreS2** (413)					

Amplification products from HBV gene regions are shown in bold with the range of fragment sizes in bp, dependent on genotype, in parentheses. Melting points (Tm) are shown for each primer. P3 and P4 are the primers described by Gunther et al. (1995) and are used to amplify the full HBV genome. Sequence in square brackets represents, non-HBV specific, primer-introduced restriction endonuclease sites which can be used to circularize the full genome product as described in the text.

products were mixed with 1/10 volume of gel loading buffer (0.25 percent bromophenol blue and 0.25 percent xylene cyanol in 30 percent glycerol) and added to the wells of a 2 to 2.5 percent agarose gel (Maniatis *et al.*, 1989). For nucleic acid detection, all agarose gels contained 0.5 µg/ml of the fluorescent dye ethidium bromide (GibcoBRL, Life Technologies Ltd, United Kingdom) and were run at constant voltage (200 to 250 Volts [V]) for one to two hours or until the darker, purple marker dye (bromophenol blue) had

progressed one-half to two-thirds of the way into the gel.

The screening method for genotyping isolates of HBV (Chapter 4) was based on size polymorphism of a PCR fragment amplified using primers specific to a section of the preS1 region and restriction fragment length polymorphism (RFLP) analysis of this, and a second PCR amplicon from the preS2 region. The PreS1 fragment (Sn2833 and Asn2973) was only 128 to 161 bp (PreS1 amplicon; Table 4) and was thus visualised by electrophoresis through a 3 percent metaphor agarose gel (FMC bioproducts) hydrated and cast according to the manufacturer's instructions. The larger PreS2 product (Sn3088 and Asn256, PreS2 amplicon; Table 4) was 413 bp and was visualised on a standard 2.5 percent agarose gel (Sigma, Figure 12).

When insufficient HBV DNA was present in a specimen, nested PCR was performed using primers Sn2833 and Asn688 (Early surface amplicon, Table 4) for the first round PCR to give a 1067-1100 bp fragment. A second PCR was then nested on the first by using 10 µl of this first round PCR reaction product as template. Reaction conditions and primers for the nested PCR were the same as for the non-nested HBV PCR reaction (see above).

Other fragments amplified for surface and core gene sequence studies included the "full surface" (Sn2833, Asn1018), "full core" (Sn1738, Asn2467), "full core and surface" (Sn1738, Asn1018) and the "full genome" (Sn1821, Asn1806) amplicons.

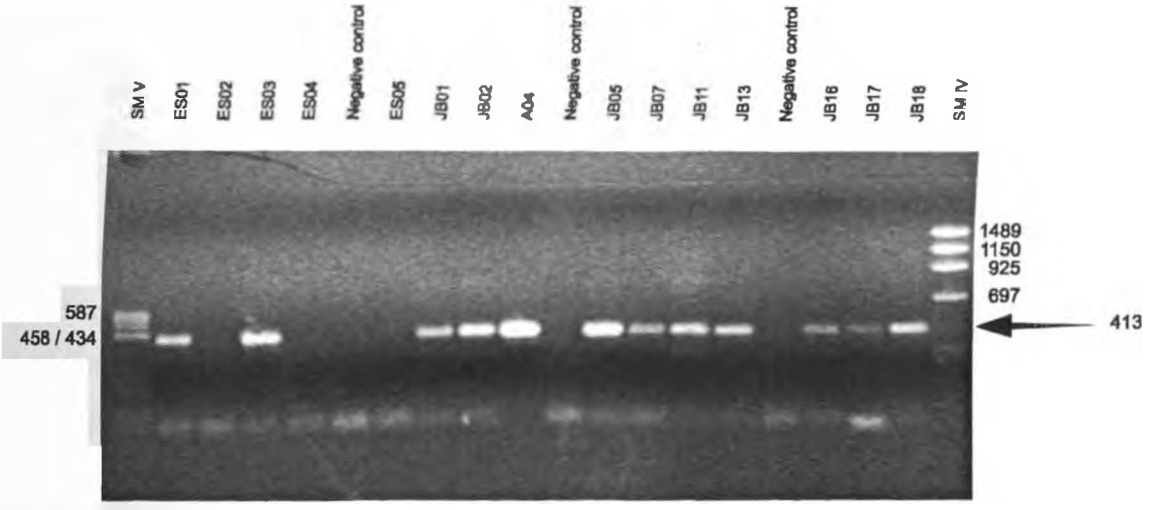


Figure 12. The HBV PCR reaction with primers Sn3088 and Asn256 results in amplification of a 413 bp fragment which includes the entire preS2 region

Fragments less than 2,000 bp used reaction conditions as described above. Larger fragments were amplified according to the protocol of Gunther *et al.* (1995) and used the Boehringer Expand Long Template System (Roche Diagnostics, Mannheim, Germany). Final reaction conditions were: 1X Roche Diagnostics (Mannheim, Germany) long PCR Reaction buffer 1 (containing 1.75 mM $MgCl_2$), 350 µM of each dNTP, 1 µM of both Sn and Asn primers and 2.5 units of Expand enzyme mix (containing thermostable *Taq* and *Pwo* DNA Polymerases (Roche Diagnostics, Mannheim, Germany). Primers were only added to the enzyme mix at the very last moment because of the proof-reading activity of *Pwo* polymerase, and reagents and reaction mixtures were kept on ice throughout to prevent non-specific amplification artifacts as described above. Ten microlitres of extracted HBV DNA in a final reaction volume of 50 µl was subject to the following cycling conditions using the Robocycler (Stratagene): 10 cycles of denaturation at 94 °C for 40 seconds, annealing at 60 °C (40 seconds) and amplification at 68 °C (2 minutes 30 seconds); 10 cycles of denaturation at 94 °C for 40 seconds, annealing at 60 °C (60 seconds) and amplification at 68 °C (5 minutes); 15 cycles of denaturation at 94 °C for 40 seconds, annealing at 60 °C (one minute and 30 seconds) and amplification at 68 °C (7 minutes); and one cycle of denaturation at 94 °C for 40 seconds, annealing at 60 °C (one minute and 30 seconds) and amplification at 68 °C (9 minutes).

3.2.3 Restriction endonuclease (RE) reactions

Ten microlitres of the appropriate PCR fragment was digested separately with 10 units of one, or more, of the following restriction endonucleases (REs; Boehringer Mannheim), *Bam HI*, *Taq I*, *Xho I* and/or *Hpa II*, using the buffers supplied and the conditions recommended by the manufacturers in a final reaction volume of 25 µl. Casein (final concentration 40 µg/ml) was added to enhance enzyme activity (Dreyer and Schulte-Holthausen, 1991). After digestion, the entire reaction volume was mixed with gel-loading dye and loaded into the wells of an agarose gel of appropriate grade and percentage. Fragments greater than 200 bp were visualised using 2 to 2.5 percent agarose D1 LE, molecular grade (supplied by Whitehead Scientific) whereas smaller fragments were run on 3 percent Metaphor agarose (FMC bioproducts) as described above.

3.2.4 Sequencing of PCR products

Prior to sequencing, PCR products were purified using either the Wizard PCR Preps DNA

purification system (Promega) or appropriate enzyme digestion as described in the following protocol. Five microlitres of PCR product was first digested with one unit of exonuclease (15 minutes at 37 °C) to degrade the unwanted primers and then heated to 85 °C for 15 minutes to denature the exonuclease enzyme. The same five microlitres of PCR product was then digested with two units of phosphatase (15 minutes at 37 °C) to remove the 5' phosphate residues from both the DNA and the dNTPs and then heated to denature the enzyme as before.

Dideoxy sequencing was performed using the Sequenase PCR-product sequencing kit (United States Biochemicals) with five microCuries of ^{35}S-dATP (DuPont NEN). Five microlitres of purified PCR product were used per reaction. In some instances the Sn or Asn primer used in the PCR reaction was also used as the sequencing primer (Table 4). More often, a relevant, dedicated sequencing-primer was used (the various alternatives are listed in Table 5). Wherever possible, sequence data was confirmed by sequencing in both forward and reverse direction using appropriate primers.

Sequencing products were resolved on 8 percent acrylamide gels (19:1 acrylamide:N, N'-

Table 5. Sn and Asn sequencing primers used for sequencing the surface and core genes

Sequence Primers	Primer sequences (5' to 3')
Sn69	CTG CAG TTC AGG AAC AGT
Sn220	GAC AAG AAT CCT CAC AAT ACC
Sn397	CAT ATT CCT CTT CAT CCT GC
Sn587	GGA AAT TGC ACC TGT ATT CCC ATC C
Sn1852	ATG TCC CAC TGT TCA AGC CTC CAA GC
Sn1957	GCC TTC TGA CTT CTT TCC
Sn2013	ATC GAG AAG CCT TAG AGT CT
Sn2187	AGA TCA GGC AAC TAT TGT GG
Asn376	TAA AAC GCC GCA GAC ACA TC
Asn2048	TTG CCT GAG TGC AGT ATG GTG AGG
Asn2297	ATA GGG GCA TTT GGT GGT CT

methlyene-bis-acrylamide [bis]) containing 8 M urea run at constant power (50 to 60 Watts) for approximately four hours or until the lighter blue, xylene cyanol, dye marker was just reaching the bottom of the gel. When Wizard preps are used for purification, gels could be run in 1X Tris-borate buffer (TBB, 90 mM Tris Borate and 2.5 mM EDTA) but following purification by enzyme digestion it was necessary to use Glycerol Tolerant buffer (GTB, 90 mM Tris pH 8, 0.5 mM EDTA and 30 mM Taurine). The gel was fixed for 15 to 30 minutes (60 minutes if a wedge gel was used) in 7.5 percent methanol and 15 percent acetic acid (10 percent methanol and 15 percent acetic for GTB) and visualised by autoradiography. Wedge gels extend the length of sequence readable on a single gel by slowing down the movement of the smaller, faster, sequence fragments which allows better separation of the larger fragments at the top of the gel. Sequence data was analysed using DNASIS 2.5 software (Hitachi).

3.3 Immunological methods

3.3.1 Lymphocyte preparation

PBMCs were isolated from either EDTA or heparinized venous blood by Ficoll-Paque (Pharmacia, Piscataway, NJ) density centrifugation.

Whole blood was centrifuged (300 g) for 10 minutes at room temperature and the plasma removed and stored at -20 °C. The top layer, which contains a mixture of plasma, buffy coat and red blood cells, was diluted 1:1 with phosphate buffered saline (PBS, pH 7.2) at room temperature and mixed well. Four volumes of the blood/PBS mixture was then slowly layered onto three volumes of Ficoll-Hypaque (Sigma) and centrifuged (900 g) for 30 minutes. Using a sterile pipette the upper layer containing the plasma and most of the platelets was discarded. The mononuclear cell layer (which is visible as a thick layer at the Ficoll interface) was transferred to another centrifuge tube and washed three times with PBS and the cells counted and their viability determined by Trypan Blue exclusion. The cells were then resuspended in complete tissue culture medium (RPMI 1640 [Flow Laboratory, UK] containing 10 mM Hepes, antibiotic [100 units/ml penicillin G and 100 μg/ml streptomycin sulphate] and indicator [0.4 percent phenol red] plus 10 percent foetal calf serum [FCS]) for immediate use, or in freezing medium (10 percent dimethyl sulphoxide [DMSO] and 20 percent FCS in complete medium) for slow freezing and storage in liquid nitrogen.

3.3.2 Antigen preparations and optimization

Branched or monomeric peptides were prepared at the Glasgow Institute of Immunology and were a kind gift from Dr. W. Carman. The branched peptides are comprised of four copies of the peptide sequence on a branched lysine core while the monomeric peptides are carboxy-terminal amides. The 17 peptides were designed so as to overlap one another and thus cover the complete core gene (Table 6).

Peptides were dissolved in water, water bubbled with ammonia vapour or acidified water (< pH 6.5) as indicated in Table 6. Parallel proliferation assays were performed initially to control for the effect of prescribed diluent. These consisted of negative controls containing no additive and positive controls containing either ammonia vapour or acidified water in the same concentrations as used to dissolve the peptides. These additives were found to have no significant effect on the counts for either control. Initially the peptides were also tested at various concentrations (0.01, 0.25, 1, 2 and 4 µg/ml) to determine the optimum concentration for peak assay performance. Results indicated that a response was elicited at the lower concentrations, although the best concentration varied from patient to patient, later assays were performed routinely at two concentrations only, namely, 0.01 and 2 µg/ml

Table 6. The position and nature of the overlapping core gene peptides

Peptide No.	Branched or Monomer	aa	Sequence	+/-	Solubility
p22-4	B	1-23	MDIDPYKEFGATVELLSFFLPSDF	4	H_2O/NH_3
p22-5	B	4-33	DPYKEFGATVELLSFFLPSDFFPSVRDLLDT	-4	H_2O/NH_3
p22-6	M	14-43	ELLSFFLPSDFFPSVRDLLDTASALYREALE	4	H_2O/NH_3
p22-7	B	24-53	FPSVRDLLDTASALYREALESPEHCSPHHT	0	H_2O
p22-8	B	34-63	ASALYREALESPEHCSPHHTALRQAILCWG	+2	H_2O
p22-9	B	44-73	SPEHCSPHHTALRQAILCWGELMTLATWVG	+2	H_2O + CH_3COOH
p22-10	M	54-83	ALRQAILCWGELMTLATWVGGNLEDPISRD	-2	H_2O/NH_3
p22-11	B	64-93	ELMTLATWVGGNLEDPISRDLVVSYVNTNM	-3	H_2O/NH_3
p22-12	B	74-103 +K3	GNLEDPISRDLVVSYVNTNMGLKFRQLLWF	0	10 % CH_3COOH or DMSO
p22-13	M	84-103	LVVSYVNTNMGLKFRQLLWFHISCLTFGRE	+3	H_2O + CH_3COOH
p22-14	M	94-113	GLKFRQLLWFHISCLTFGRETVIEYLVSFG	+2	H_2O + CH_3COOH
p22-15	M	104-123	HISCLTFGRETVIEYLVSFGVWIRTPPAYR	+2	H_2O + CH_3COOH
p22-16	M	114-133	TVIEYLVSFGVWIRTPPAYRPPNAPILSTL	+1	H_2O + CH_3COOH
p22-17	B	124-153	VWIRTPPAYRPPNAPILSTLPETTVVRRRG	+4	H_2O
p22-18	B	134-163	PPNAPILSTLPETTVVRRRGRSPRRRTPSP	+6	H_2O
p22-19	B	144-173	PETTVVRRRGRSPRRRTPSPRRRRSQSPRR	+12	H_2O
p22-20	B	154-183	RSPRRRTPSPRRRRSQSPRRRRSQSRESQC	+12	H_2O

(Appendix E, Table 27a).

3.3.3 Cell culture conditions

Cell cultures for measuring the proliferative responses were performed in 96-well, round-bottomed microtitre plates containing 1 X 10^5 cells/well in a total of 200 µl RPMI 1640 with 10 percent FCS. The peptide analogues were diluted as described above according to the instructions from the Glasgow Institute of Immunology. Each peptide was cultured in quadruplicate at two concentrations 0.01 and 2 µg/ml. The cultures were incubated in an atmosphere containing five percent CO_2 at 37 °C for four to five days (see below).

3.3.4 Proliferation assay

For each patient specimen, four replicates of each peptide at two concentrations (2 µg/ml and 0.01 µg/ml determined empirically as described above) were added in 50 µl to the wells of a 96-well plate together with four positive control wells containing 10 µg/ml phytohaemagglutinin (PHA), a polyclonal B cell activator (PBA), and four negative control wells containing complete medium. The final concentration of 10 µg/ml PHA which was used routinely in the assays was chosen by comparing the stimulation indices obtained from a range of concentrations [2.5, 5, 10, and 20 µg/ml; Appendix E, Table 27(a)]. To this was added 20 µl of FCS (10 percent in final volume) and 130 µl complete medium containing 2 X 10^5 resuspended PBMCs to give a final volume of 200 µl per well. The plate was then incubated in a humidified chamber at 37 °C in five percent CO_2 for 90 hours.

After 90 hours, each well was pulsed with 0.25 µCi tritiated thymidine (^{3}H-thymidine; methyl-^{3}H 40 to 60 Ci/mmol, Amersham) and incubated for a further 6 to 18 hours. Cells were then harvested onto glass fibre filter disks (Skatron, Sterling, USA) and incorporated ^{3}H-thymidine measured by counting each disk in 10 ml of scintillant (Ultima Gold XR, Packard Bioscience, The Netherlands) in an LKB 1205 β liquid scintillation counter.

The stimulation index, or SI, is defined as the mean counts per minute of ^{3}H-thymidine uptake obtained in the presence of antigen to that obtained in the absence of antigen. An SI value of greater than 2.5 was regarded as positive.

3.3.5 Class II (HLA-DR) typing

QIAamp columns were used to extract total genomic DNA (as described in paragraph 3.2.1 above) from PBMCs isolated from each of the patients whose cells were tested in the proliferation assay. For a final recommended DNA concentration of 50 nanograms (ng) per microlitre, the DNA was eluted in one-half to one millilitre (ml) of buffer for every 200 μl of buffy coat or 0.6 to 0.8 x 10^7 cells applied to the QIAamp column. The HLA-DR class II type was then determined for all patients using sequence-specific PCR (SSP) and primers supplied with the low-resolution-SSP HLA-DR typing kit (Dynal, Norway) in a series of 24 HLA-DR specific PCR reactions according to the manufacturer's instructions. HLA type was deduced from the pattern of amplified fragments of specific size in each reaction tube visualised by agarose gel electrophoresis (D1 LE, molecular grade agarose supplied by Whitehead Scientific, Figure 13) and interpreted using the appropriate specificity tables supplied with the kit.

Results of the more commonly occuring HLA-DR types, DRB1*03 and DRB1*11, as well as two of the rarer types, DRB1*08/12 and DRB1*16 were checked and refined using high resolution-SSP HLA typing kits. Figure 14 shows the gel obtained with specimen HB05 and the DRB1*03 kit. Whereas the low resolution PCR series could not distinguish between the

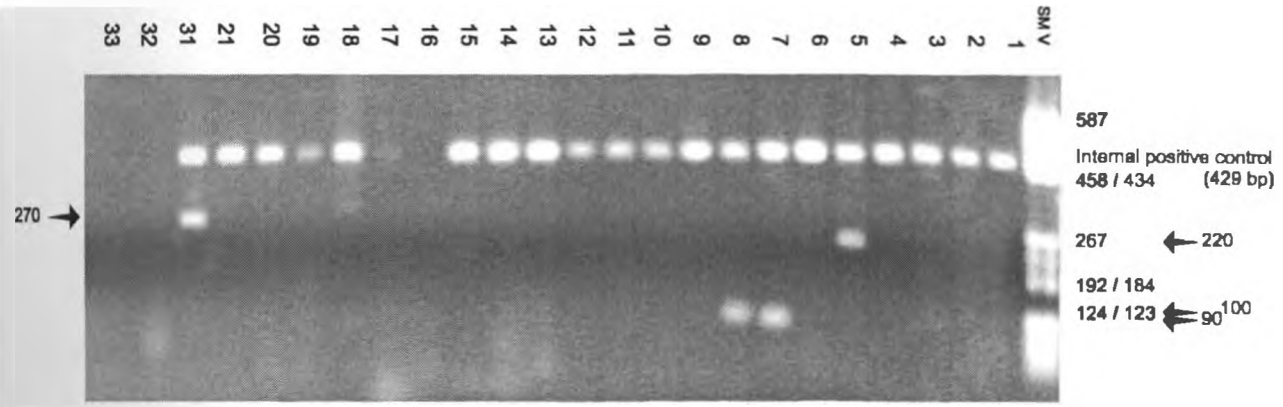


Figure 13. HLA-DR specific bands for patient ES06 are present in lanes 5 (220 bp), 7 (90 bp), 8 (100 bp) and 31 (270 bp). The presence of bands in lanes 5 and 7 but not lane 6 indicates the presence of the allele DRB1*0302 or *0303, the DR18 alleles, eliminating several other possible DR3 alleles. This is confirmed by the presence of the DRB3 allele as indicated by the presence of the band in lane 31. The presence of a band in lane 8 is indicative of the DRB1*04 group of alleles but the specific allele present cannot be determined more accurately with the low resolution kit

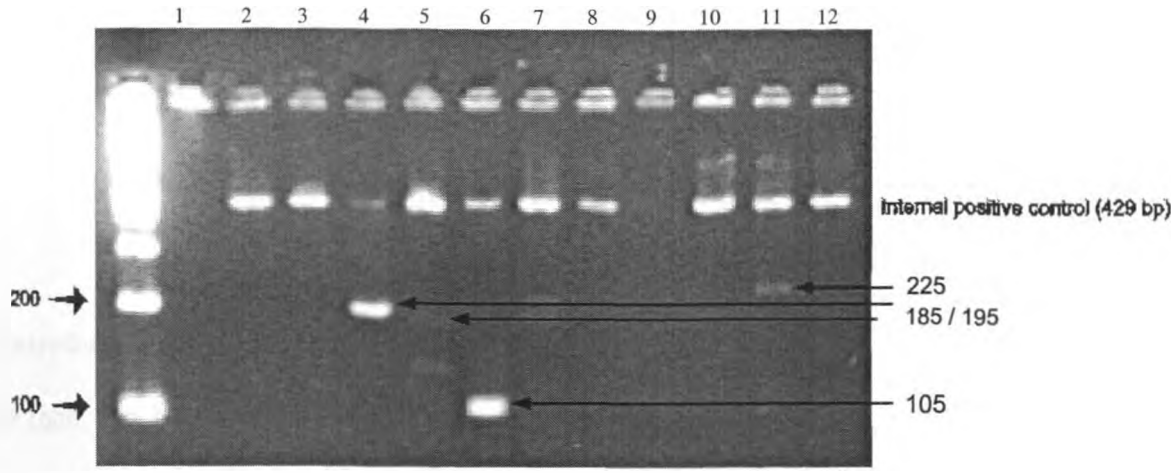


Figure 14. High resolution HLA-DR3 specific bands for patient HB05 present in lanes 4, 5, 6 and 11 not only confirm that the patient has one of the DR3 group of alleles but also pinpoints that the specific allele present is DRB1*0311

alleles DRB1*0301-0317, the positive bands in lanes four (195 bp), 5 (185 bp), 6 (105 bp) and 11 (225 bp) of the high resolution PCR series (Figure 14) occur together only in the allele DRB1*0311.

CHAPTER 4

THE EPIDEMIOLOGY OF HBV INFECTION

4.1 Introduction

More than two billion people worldwide have evidence of past or present HBV infection (WHO update, 1996) and it is estimated that 70 percent of the world's population live in areas of moderate (carrier rate greater than two percent) to high (carrier rate greater than 8 percent) HBV endemicity (Figure 15; Maynard, 1990). Moreover, carrier rates can exceed 20 percent in the main hyper-endemic areas of East Asia and sub-Saharan Africa. Asia supports 75 percent of the more than 385 million HBV chronic carriers worldwide (Kiire, 1993) and sub-Saharan Africa has 50 million carriers in a total population of 415 million people (Kew, 1998).

The infection is acquired early in life and transmission may be from mother to infant during the perinatal/neonatal period or by horizontal spread via other modes of transmission. East Asia has been found to have higher prevalences of HBeAg positive mothers and thus a greater risk of perinatal transmission than sub-Saharan Africa, although these differences are not statistically significant (Edmunds *et al.*, 1996). Although vertical transmission is not considered a major mechanism of HBV infection in South Africa, a recent study in our laboratory found a high prevalence of HBsAg (8.1 to 8.9 percent) in children less than one year old (Vardas *et al.*, 1999).

The primary causes of death from HBV are cirrhosis and HCC and 25 to 30 percent of carriers who survive for 30 years or more will subsequently die of the sequelae of their infection. HBV causes as much as 80 percent of the HCC in sub-Saharan Africa and most parts of the Far East (Maynard *et al.*, 1990). Early-onset carriers are at very high risk of developing HCC (lifetime relative risk > 100, Beasley and Hwang, 1984) making HBV a preventable environmental carcinogen second only to cigarette smoking (Kew, 1998).

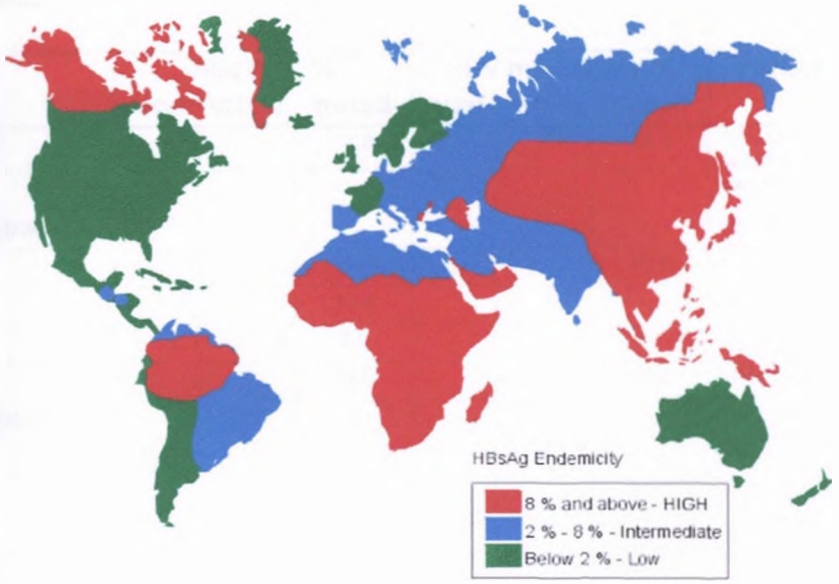


Figure 15. Prevalence of hepatitis B virus infection worldwide

4.1.1 Epidemiology of HBV infection in southern Africa

Epidemiological studies of HBV infection in sub-Saharan Africa are numerous (Table 7) although population groups studied vary greatly in their distribution (rural, urban or a mixture of rural and urban), composition (age and sex stratification) and size. Despite these limitations and with the exception of a few urban areas, these studies confirm that HBV is highly endemic in sub-Saharan Africa.

4.1.1.1 Regional prevalence studies in southern Africa

Regional studies undertaken in southern Africa reveal hyper-endemic foci and regional differences in the prevalence of both HBsAg and other markers of HBV infection. Although the South African national average of HBsAg prevalence is 9.9 percent (Dusheiko *et al.*,1989), within the high endemicity range, the average prevalence reported for the Highveld provinces of the Free State and the former Transvaal, which now comprises the North and North West Provinces, Mpumalanga and Gauteng, have intermediate endemicity for HBV (HBsAg prevalence two to 8 percent; Figure 16; Dusheiko *et al.*,1989). Average figures show the Eastern Cape to have the highest prevalence of HBsAg in South Africa (12.3 percent) with hyper-endemic pockets, exceeding 15 percent (Dusheiko *et al.*,1989), in the Transkei and Ciskei. Kwazulu-Natal has an average prevalence between these two extremes (11.3 percent) but in rural areas of this province, particularly near the Mozambique border, prevalences as high as 18.5 percent have been reported (Abdool Karim *et al.*, 1988).

Table 7. Prevalence of HBV serological markers in sub-Saharan Africa

Country	HBsAg+, %			HBV marker (+), %			Patients studied	Ref
	rural	urban	mixed	rural	urban	mixed		
Angola			12.8			79.1	201 adult volunteers (hospital patients)	1
Botswana	11.6						male miners	2
Botswana central	17.3						male miners	2
Burundi			11.0			76.0	adult population	3
CAR			15.4				hospital patients	4
Congo			8.0				blood donors	5
Ethiopia			11.0			79.0	adult population	3
semi-urban Gabon		13.8					cluster sampling of 30,000 adults	6
Gabon	19.0						303 villagers	7
Gambia			10.0			90.0	adult population	3
Ghana	20.9			74.7			1,385 sample, all ages	8
Ivory Coast			17.5					9
Kenya			11.4			56.2	adult population	3
Lesotho	8.1						male miners	2
Maseru, Lesotho		8.8					male miners	2
Liberia			16.2				infants (0-4 y)	10
Madagascar	26.0			65.4			cluster sampling of 30 communities	11
urban Madagascar		5.3			34.2		cluster sampling of 30 communities	11
Malawi	7.5						male miners	2
Mali			11.3			97.6	adult population	3
Mauritania			20.3				349 blood donors	12
Mozambique			14.6			75.0	adult population	3
Mozambique			10.5				male miners	2
Namibia			14.0			87.5	adult population	3
Niger			29.8				group of control hospital patients	13
Nigeria			10.0			72.5	adult population	3
Senegal			11.8			91.0	adult population	3
Sierra Leone		16.6					66 urban primary school children	14
South Africa (SA)			9.6			76.0	adult population	3
Sudan	18.7			63.9			851 villagers (1-89 y)	15
Swaziland	13.8						male miners	2
Siteki, Swaziland	20.3						male miners	2
Tanzania			11.0				300 blood donors	16
Zaire			20.6			78.9	adult population	3
Zambia	6.0			55.0			620 villagers (0-89 y)	17
Zimbabwe	15.3	15.7		90.9	92.0		3,394 men & women (10-61 y)	18

1 Steele and Bos, 1996
2 Dusheiko *et al.*, 1989
3 Kiire, 1993
4 Meunier *et al.*, 1984
5 Ibara *et al.*, 1991
6 Bertherat *et al.*, 1998
7 Richard-Lenoble *et al.*, 1995
8 Martinson *et al.*, 1998
9 Lohoues-Kouacou *et al.*, 1998
10 Prince *et al.*, 1981
11 Boisier *et al.*, 1996
12 Lo *et al.*, 1999
13 Cenac *et al.*, 1995
14 Hodges *et al.*, 1998
15 Hyams *et al.*, 1989
16 Matee *et al.*, 1999
17 Tabor *et al.*, 1985
18 Tswana *et al.*, 1996

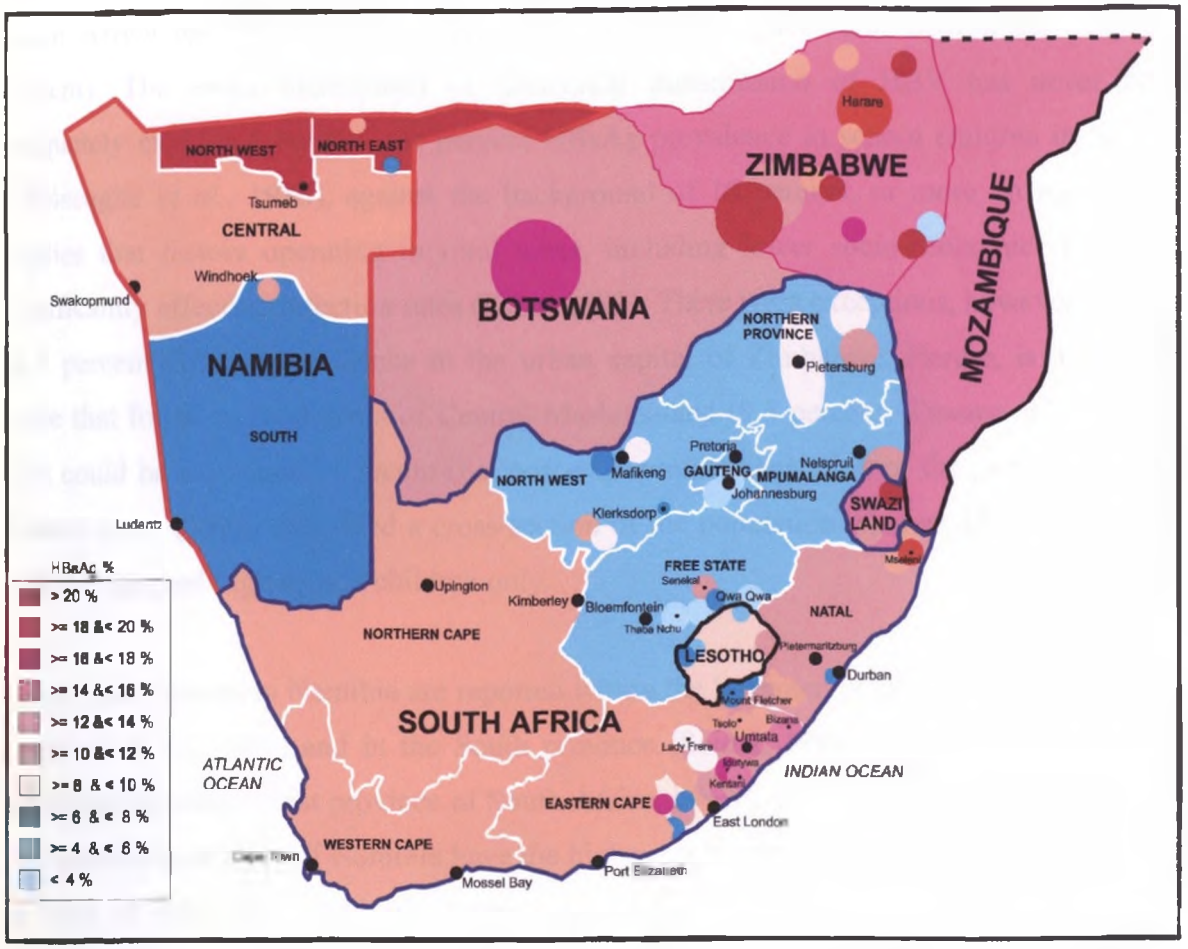


Figure 16. Percentage prevalence of HBsAg in various towns and regions of southern Africa (Seidel, 1997; Botha *et al.*, 1984; Dusheiko *et al.*, 1989; Abdool Kariem *et al.*, 1988; Vardas *et al.*, 1999; Tswana *et al.*, 1996)

Most of the regional data for South Africa, Swaziland, Lesotho and Botswana comes from a survey of male mine workers (Dusheiko *et al.*, 1989). Since HBV infection is gender dependent, these figures could be exaggerated. They do, however, indicate the trends within the population. Of interest, results from a recent survey by Vardas *et al.* (1999), a community-based, cross-sectional, age-stratified study of a sample of children from both urban and rural regions of the Mdantsane district (which abuts the East London district in the North and East), agreed well with those of Dusheiko *et al.* (1989) from the same district. Other than the average former Cape province prevalence of 12.3 percent, no information is available for the North and Western Cape. The Northern Cape Province, the largest of the 9 provinces of South Africa, has a very low population density due to the low rainfall in this region.

In general, urban areas were found to have a much lower endemicity than rural areas. For example, Carletonville (3.4 percent) and Soweto (one percent) in the Gauteng province of

South Africa have a much lower prevalence of HBsAg than rural Bophuthatswana (8.9 percent). The exact mechanism of horizontal transmission of HBV has never been adequately explained but the one percent HBsAg prevalence in school children in Soweto (Dibisceglie *et al.*, 1986), against the background of 10 percent, or more, in rural areas, implies that factors operating in rural areas, including lower socio-economic status, are significantly affecting infection rates (Kiire, 1993). There were exceptions, however, and the 18.8 percent HBsAg prevalence in the urban capital of Zimbabwe, Harare, is more than twice that found in rural areas of Central Mashonaland (9.3 percent; Tswana *et al.*, 1996). This could be explained by recent (i.e. post-exposure) urbanization of the population since Tswana *et al.* (1996) examined a cross-section of the population whereas Dibisceglie *et al.* (1986) examined high school children only.

Lowest prevalences in Namibia are reported within the bushmen of the North East province (Aspinall *et al.*, 1994) and in the South province (Seidel, 1997). This province, like the neighbouring North West province of South Africa, is very dry and sparsely populated. The two northern provinces of Namibia have the highest prevalence of past and present exposure to HBV of the entire southern African region and with Zimbabwe, Mozambique and Swaziland form the so-called "Eastern corridor". Overall, 14,8 percent of Namibian first time blood donors had markers of recent exposure to hepatitis B or were chronic carriers, and the highest prevalence, 23 percent, was observed in the North East province (Seidel, 1997).

4.1.2 Serotypes of HBV

HBV strains are divided into 9 major serotypes, *adw2*, *adw4 q* , *ayw1*, *ayw2*, *ayw3*, *ayw4*, *adr q+*, *adr q* and *ayr* (Courouce-Pauty *et al.*, 1976; 1978). These divisions are based on a common *a* determinant and two mutually exclusive *d/y* and *r/w* determinants of the major envelope protein and have been widely used to differentiate the strains of HBV in clinical, virological and epidemiological studies (Norder *et al.*, 1993). Serotypes have been shown to be geographically distributed (Courouce-Pauty *et al.*, 1983). As can be seen in Figure 17 serotype *adr* is confined to the East, *ayw2* predominates in the Mediterranean and *adw4* is only found in the Pacific region. Serotype *adw2* predominates in Northern Europe and sub-Saharan Africa, but its prevalence declines as one moves from East Africa towards Central and West Africa accompanied by a corresponding increase in the unique African serotype,

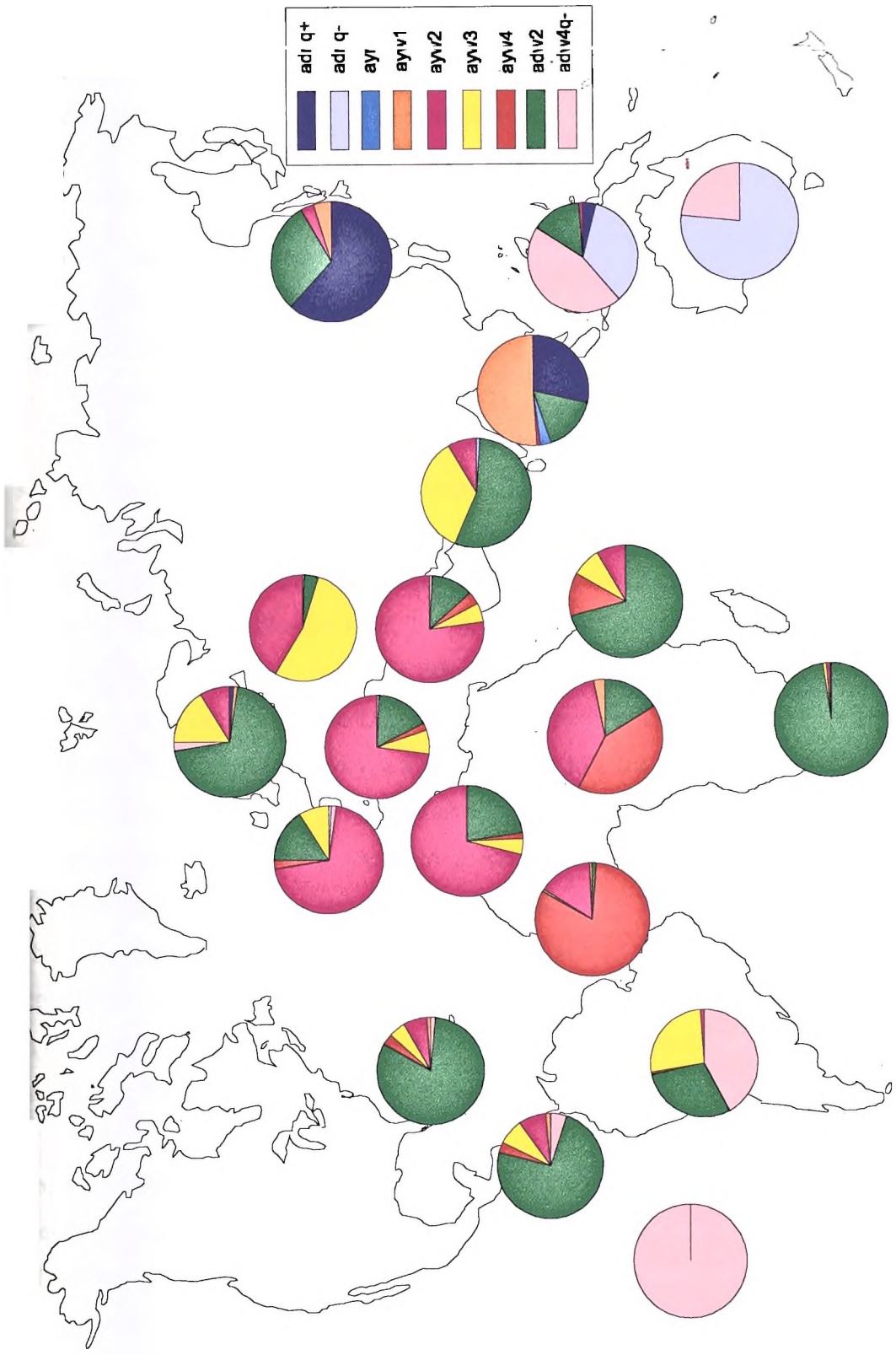


Figure 17. World distribution of HBsAg serotypes (Couroucé-Pauty *et al.*, 1983)

ayw4 (Magnius and Norder, 1995). Serotype *adw2* was previously reported to constitute 97.7 percent of HBV serotypes in South Africa, with the remaining 2.3 percent being *ayw*. No figures are available for Namibia although West Africa is reported to have a 98.8 percent incidence of serotype *ayw* of which 82.3 percent is reported to be of the unique African serotype *ayw4*. The remaining 1.2 percent is of type *adw* (Courouce-Pauty *et al.*, 1983).

4.1.3 Genotypes of HBV

Alignment and comparison of sequence data from various parts of the world originally divided HBV strains into four genotypic groups, A to D (Okamoto *et al.*, 1988) based on pairwise differences in nucleotide sequences of 8.0 percent or more. As further sequence data from other parts of the world became available, this classification was extended to include another two groups, genotypes E and F (Norder *et al.*, 1992). Last year a putative seventh genotype, genotype G, was proposed (Stuyver *et al.*, 2000).

Dendrograms drawn from available sequence data reveal a characteristic global distribution among the genotypes (Norder *et al*, 1993). Some genotypes, such as genotype A, could be associated with the West, whereas others such as genotypes B and C were primarily confined to the East. Genotype D is fairly widespread but is the predominant group in the Mediterranean countries. Genotype E is localised to parts of East, Central and West sub-Saharan Africa. Genotype F, regarded as indigenous to the Amerindian populations in the Americas, has recently been found to predominate in Central American populations of Hispanic origin as well, but remains specific to the New World (Arauz-Ruiz *et al.*, 1997a). The exact distribution of genotype G globally is not yet known. The original isolates represented 11 percent of a cohort of 121 plasma specimens collected from chronic carriers in France and the United States (Stuyver *et al.*, 2000). A recent study of 540 sera from HBV patients in Japan did not detect any genotype G (Kato *et al.*, 2001).

Sequence studies have also shown serotypic differences to be based on discrete and limited sequence changes, and that when isolates of known serotype are sequenced, no genetic relatedness is evident (Figure 18; Yang *et al.*, 1995a). Several of the genotypic groups are serotypically heterogenous: genotype C contains serotypes *adw2*, *adr* and *ayr*; both genotype A and B contain serotype *adw2* and *ayw1*; genotype D contains only the *yw*

serotypes *ayw2* to *ayw4*. Some serotypes are spread across the genotypic groups (*adw2* occurs in the genotypic groups A, B, C and F [Blitz et al., 1998] while *ayw1* occurs in genotypes A and B). Serotype *adw2* clusters into genotype A in the West but B and C in the East; *ayw1* also occurs in genotype A in the West, but is confined to genotype B in the East (Figure 18; Magnius and Norder, 1995). The 11 specimens of genotype G studied to date were predicted to be serotype *adw2*.

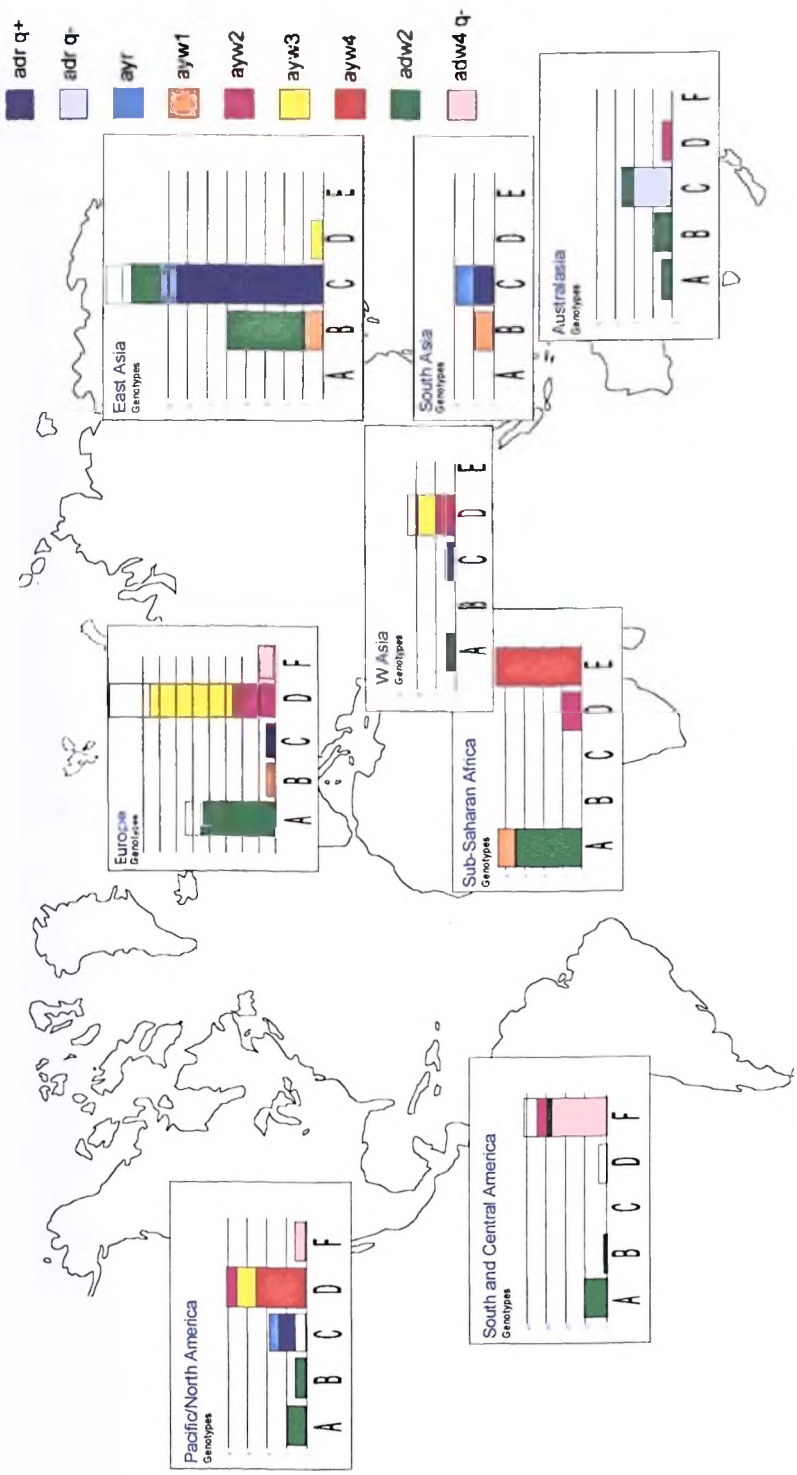


Figure 18. World distribution of HBV genotypes/serotypes
Proportions of serotypes within the various genotypes in various parts of the world were estimated from GenBank data within my database of specimens used for comparison in the various studies reported in this thesis.

4.1.4 Molecular epidemiological studies and strategies

Although the 9 basic serotypes do not reflect the underlying heterogeneity of the HBV genome (Okamoto *et al.*, 1988), many of the molecular epidemiological studies to date have been based on these partitions. While genotypes do reflect genetic relatedness between viral strains little information is available regarding the genotypes circulating in sub-Saharan Africa. For this reason I sought a screening assay to determine genotype variation among local HBV-infected patient sera. My aim was to find an accurate method to partition sera into genotypes which also identified exceptional/unique specimens for further sequence analysis. It was important that the final screening assay did not require a high level of skill in molecular techniques so that it could be performed in a fairly basic molecular diagnostic laboratory, since several requests from neighbouring countries for such an assay had been made.

A number of different methods which adhere to these criteria have been used to determine genotype and these include: PCR with type-specific primers, DNA single-stranded conformational polymorphism, reverse phase hybridisation and restriction fragment length and size polymorphism (Table 8). I considered the merit and adaptability of the strategies available at the start of the study and then continued to adapt my methodology as new information, applicable to my purpose, was reported. The method finally adopted used restriction fragment length polymorphisms (RFLP) in the preS region and a preS1 gene size polymorphism (Shih *et al.*, 1991; Okamoto *et al.*, 1988) to assess the heterogeneity of HBV.

Recently a serological determination of genotype, which uses an EIA assay and monoclonal antibodies raised against genotype-specific epitopes in the preS2-region product, has been developed (Usuda *et al.*, 1999). In collaboration with the authors, this technique was used to type selected southern African isolates.

4.2 Materials and Methods

4.2.1 Patients

One hundred and forty-one HBsAg-positive specimens from the groups described in paragraph 3.1 of Chapter 3 were successfully screened for genotype using RFLP and size polymorphism of amplicons and/or HBV genotype-specific EIA and/or sequence analysis.

Table 8. Different published methods of HBV serotyping and/or genotyping

Study	Types discerned	Screening method
Okamoto *et al.*, 1988	restricted genotypes	similarity between 18 full genomes
Norder *et al.*, 1990	serotypes	type-specific primers
Shih *et al.*, 1991	serotypes	RFLP + size polymorphism
Sastrosoewignjo *et al.*, 1991	restricted genotypes	similarity between HBsAg sequences
Norder *et al.*, 1992	restricted genotypes	phylogenetic analysis of HBsAg sequence
Uy *et al.*, 1992	restricted genotypes	phylogenetic analysis of preS1 sequence
Li *et al.*, 1993	genotypes	size polymorphism
Norder *et al.*, 1993	genotypes	phylogenetic analysis of HBsAg sequence
Repp *et al.*, 1993	serotypes	multiplex PCR with type-specific primers
Kidd-Ljunggren *et al.*, 1994	genotypes	phylogenetic analysis of preS2 sequence
Niel *et al.*, 1994	serotypes	RFLP
Ohba *et al.*, 1995	restricted genotypes	phylogenetic analysis of full genome
Moraes *et al.*, 1996	genotypes	phylogenetic analysis of preS/S sequence
Zhang *et al.*, 1996	restricted genotypes	size polymorphism
Hardie *et al.*, 1996	strains	DNA single stranded conformational polymorphism
Arauz-Ruiz *et al.*, 1997b	genotypes	HBsAg sequence comparison
Bowyer *et al.*, 1997	genotypes	phylogenetic analysis of preS2/S sequence
Lindh *et al.*, 1997	genotypes	RFLP
Mizokami *et al.*, 1997	restricted genotypes	phylogenetic analysis of full genome
Telenta *et al.*, 1997	genotypes	phylogenetic analysis of partial HBsAg sequence
Mbayed *et al.*, 1998	genotypes	phylogenetic analysis of partial HBsAg sequence
Blitz *et al.*, 1998	serotypes/genotypes	reverse-phase hybridisation
Lindh *et al.*, 1998	genotypes	RFLP
Mizokami *et al.*, 1999	genotypes	RFLP

Of the 93 South African specimens typed, 56 came from the urban areas of Gauteng (paragraph 3.1.1.1) and the remaining 37 were from the rural regions of the Eastern Cape (paragraph 3.1.1.2; Vardas, 1999). All of the rural specimens and 38 of the urban specimens were genotyped by size polymorphism and RFLP analysis, 25 of the urban specimens were also sequenced and a further 12 were genotyped by sequence and phylogenetic analysis alone. Twelve specimens were genotyped by HBV genotype-specific EIA and of these 6 were also genotyped by size polymorphism and RFLP analysis and five of these were also sequenced.

The 49 genotyped specimens from the northern provinces of rural Namibia (paragraph 3.1.2) included 16 specimens from the pilot HBV prevalence study of the northern provinces. Thirteen of these were typed by size and RFLP polymorphism and three by genotype-specific EIA. Eighteen of the 25 typed specimens from the 1984 Ovamboland study were typed by size and RFLP polymorphism, 9 of these were also sequenced and 6

were also typed by genotype-specific EIA. The remaining 7 specimens were typed by genotype-specific EIA alone. Finally, all 8 Nambian patients from my proliferation assay study were successfully typed by HBV genotype-specific EIA (N231-N238).

4.2.2 Genotype screening

4.2.2.1 Rationale for RFLP and size polymorphism analysis

Genotype D isolates can be differentiated from isolates from all other genotypes on the basis of the size of their preS1 region which has a 33 bp deletion. The preS1 region of genotype E isolates has a 3 bp deletion but this is not sufficient to distinguish them unambiguously from isolates of genotype A, B, C or F. Fmy GenBank specimens, two from genotype C and two from genotype A, reportedly have deletions varying from 15 to 21 bp which would result in the amplification of an anomalous preS1 fragment.

My final selection of restriction endonucleases (REs) for strain mapping was based on the consistency of the presence or absence of RE sites within the preS1 and preS2 PCR fragments (referred to as PS1 and PS2, respectively) and, in particular, the specificity of the pattern of these sites within the different strains of HBV as predicted using sequence data from GenBank and the literature.

Since the screening assay is based on limited nucleotide change, I continually checked newly released GenBank sequences to re-confirm my assumptions and also verified the results obtained in the screening assay by performing parallel phylogenetic analysis on sequence data from 46 local specimens.

A 3 to 33 bp (dependent upon genotype) deletion in the preS1 region of the HBV genome together with 6 *Bam HI*, *Taq I*, *HpaII* or *Xho I* sites (Figure 19) reliably partitioned isolated DNA into the respective HBV genotypes (Table 9). The *Bam HI*(2943) site, which is present in isolates from genotypes D, E and F, is particularly useful when used in conjunction with this size polymorphism. A *Taq I* site within the deletion (site 2877) was found to be consistently present in genotype B and in all genotype C isolates containing the *q* determinant (i.e. not Cq). Genotype E also has this *Taq I* site, although it is displaced slightly, occurring at nucleotide 2879. Genotypes D to F can thus be identified on the basis of PS1 alone (Table 9).

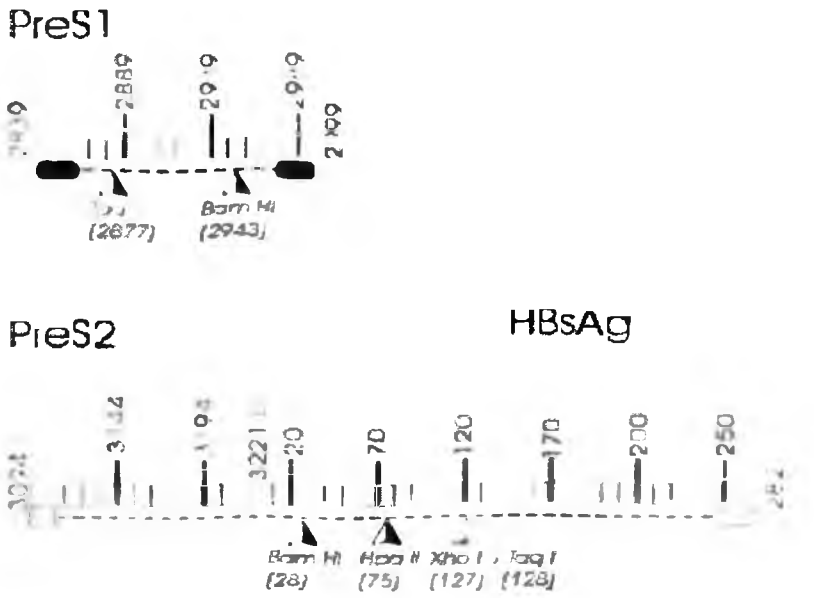


Figure 19. Restriction endonuclease sites used in the RFLP genotype screening assay

In my final regimen, having distinguished genotypes D, E and F from the rest, I then performed a *Bam HI* digestion on PS2 to distinguish genotype A, subgroup A, specimens from those that remained. A PS1 *Taq I* site but no PS2 *Xho I* site was unique to genotype B, and genotype C specimens could be separated from genotypes A and B on the basis of the

Table 9. Size and restriction fragment length polymorphisms used to partition isolates into the 6 major genotypes and two subtypes

	SIZE PS1	PS1 + Bam HI (2943)	PS1 + Taq I (2877)	PS2 + Bam HI (28)	PS2 + Hpa II (75)	PS2 + Taq I (128)	PS2 + Xho I (127)
D	128	78/50	-	-	-	+/-	+/-
E	158	108/ 50	111/ 47	162/ 251	202/ 211	262/ 151	261/ 152
F	161	111/ 50	-	162/ 251	-	262/ 151	-
A	161	-	-	162/ 251	-	-	-
B	161	-	116/ 45	-	-	262/ 151	-
A'	161	-	-	-	-	262/ 151	+/-
C	161	-	116/ 45	-	202/ 211	262/ 151	261/ 152
Cq-	161	-	-	+/-	202/ 211	262/ 151	261/ 152

The size of PS1 (the preS1 fragment) which distinguishes genotype D from all other genotypes is shown in column 2. The columns to the right of this show the relevant fragment sizes of either PS1 or PS2 (the preS2 fragment) after the digestion indicated in the column header. Alternatively '-' indicates that the site is absent while '+/-' indicates that some, but not all, specimens of the genotype in question have this site

Hpa II(75) site. The *Taq I*(2877) site could further distinguish the *q* and q^+ subtypes of genotype C. Subgroup A' specimens lacked both *Bam HI* sites and the *Taq I*(2877) and *Hpa II*(75) sites but could be positively separated from the rest of genotype A by digestion with *Taq I*(128).

Table 9 summarizes the occurrence and size of the various digestion products of PS1 and PS2 and illustrates the unique patterns obtained for each of the strains of HBV. The order of the digests can be varied according to the expected occurrence of the genotypes in the population being studied and this reduces the total number of digests required per specimen.

4.2.2.2 PCR amplification of HBV DNA

Having chosen suitable restriction enzyme sites and size polymorphisms for the screening assay, I then chose appropriate primers for amplification of PS1 and PS2 (DNASIS, Hitachi software). My original selection used Sn2833/Asn3088 to give a 239 to 273 bp amplicon in the preS1 region and Sn3088/Asn256 to give a 413 bp fragment in the preS2 region. The preS1 Asn primer, Asn3088, was later replaced with Asn2973 (Li *et al.*, 1993), which gave a smaller (128-161 bp) fragment, since subsequent sequence analysis revealed that Asn3088 (5' TATGCCCTGAGCCTGAGG 3') frequently bound at nucleotide 3203 which resulted in a larger than expected PS1 amplicon. Further experimental detail of the PCR, primers, digestions and agarose gel electrophoresis are given in paragraphs 3.2.2 and 3.2.3.

4.2.2.3 Rationale for HBV genotype-specific EIA

Variation in the nucleotide sequence of the preS2 region results in a corresponding variation in the preS2 region product and this can also be used to distinguish HBV genotype (Usuda *et al.*, 1999). Five monoclonal antibodies (mAbs; *b*, *m*, *k*, *s* and *u*) were raised against the preS2-region products of HBV isolates of various genotypes. Epitope *b* (amino acid 20-23 of the preS2 product) was found to be expressed commonly by all genotypes. Expression of the other four epitopes (*m* [33-52], *k* [37-46], *s* [14-21] and *u* [6-27]) recognized by the mAbs partitioned specimens into four HBV genotypes (A, B, C and F) and a fifth partition in which genotypes D and E were inseparable. To distinguish genotype D from genotype E a further two mAbs (*f* [7-15] and *g* [42-48]) were raised. Expression of both these epitopes was characteristic of genotype E but not genotype D. These two genotypes could thus be

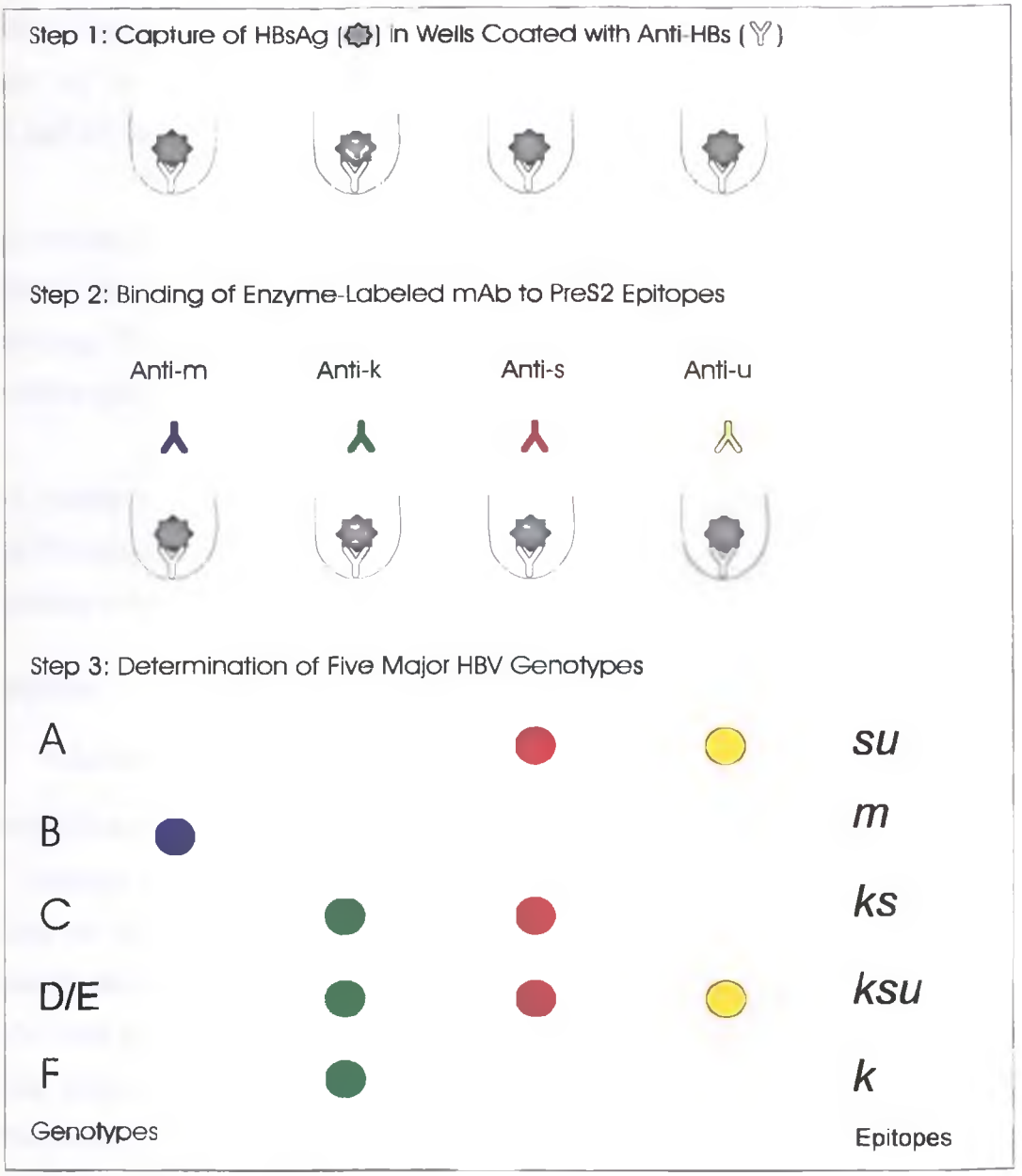


Figure 20. Serological partitioning of isolates into genotypes A, B, C and F as well as a fifth partition made up of specimens from both genotype D and E using the HBV genotype-specific EIA

distinguished using an additional sandwich EIA in which the wells were coated with the mAb to epitope *g* and labelled mAb to epitope *f* was used for detection (Usuda *et al.*, 2000).

Figure 20 shows the three steps of the initial HBV genotype-specific EIA. In the first step, HBsAg in the test serum is captured by wells of a microplate coated with mAb to the common determinant *a* of HBsAg. The plate is then washed. In the second step, each of the four wells (for a test) receives enzyme-labelled mAb to one of the epitopes *m*, *k*, *s* or *u*. The plate is again washed, and the colour developed in the wells. In the third and last step,

genotypes are determined by the combination of preS2 epitopes (serotypes) recognized by mAbs: "*su*" for genotype A, "*m*" for genotype B, "***ks***" for genotype C, "***ksu***" for genotype D or E, and "*k*" for genotype F.

Since epitope *b* is expressed on the preS2-region products of all genotypes it was used to determine the presence of the preS2-region product as a prerequisite to the serological genotyping. This was found to be necessary because the preS2-region product is easily digested in specimens contaminated with bacteria or proteases.

I also examined the preS2 sequence of 166 sequences from GenBank, which included 26 South African specimens, to determine the incidence of variation, if any, in the regions of the genome involved in the three dimensional (3D) variation of the preS2 product.

4.3 Results

4.3.1 Size polymorphism results

Genotype D specimens were clearly differentiated from all other genotypes by their smaller size. Genotype E PS1 amplicons could not be distinguished on the basis of size as their 3 bp deletion was insufficient for accurate identification. None of the laboratory specimens had anomalous deletions as reported for four GenBank specimens. However, on four occasions a double band was observed on the gel, indicating a mixture of genotype within certain patients (Figure 21, lane 4 and lane 19). This could also have been the result of contamination although the aberrant results were isolated and not clustered. Although all of the HBsAg- and HBeAg-positive specimens from northern Namibia (JB numbers) gave PS1 amplicons of the expected size, more than half (18/34) of the specimens from the pilot study from the northern regions gave a larger than expected PS1 amplicon, either on its own or together with the expected amplicon (Figure 22).

4.3.2 South African genotyping results

More than half of genotype A specimens from South Africa were found to have a digestion pattern similar to the genotype A sequence published by Estacio *et al.* (1988). This was so common that I called this subset of genotype A subgroup A' (Bowyer *et al.*, 1997) and its characterization is reported in Chapter 5 of this dissertation. Following the characterization of the subtype A' of genotype A, the size of the PS1 amplicon together with the presence or

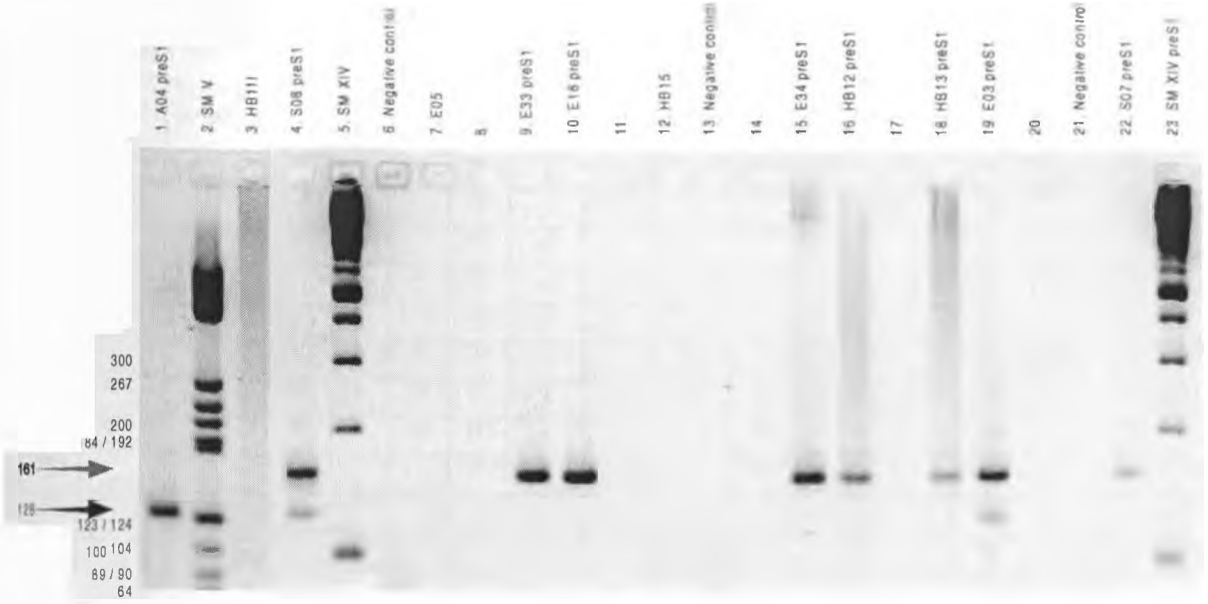


Figure 21. PS1 amplicons on 3 percent Metaphor agarose showing the clear distinction between the 128 bp amplicon of genotype D and the larger (158-161 bp) amplicons of all other genotypes. Two specimens (lane 4 and lane 19) showed mixed genotype with bands at both 128 and 161 bp

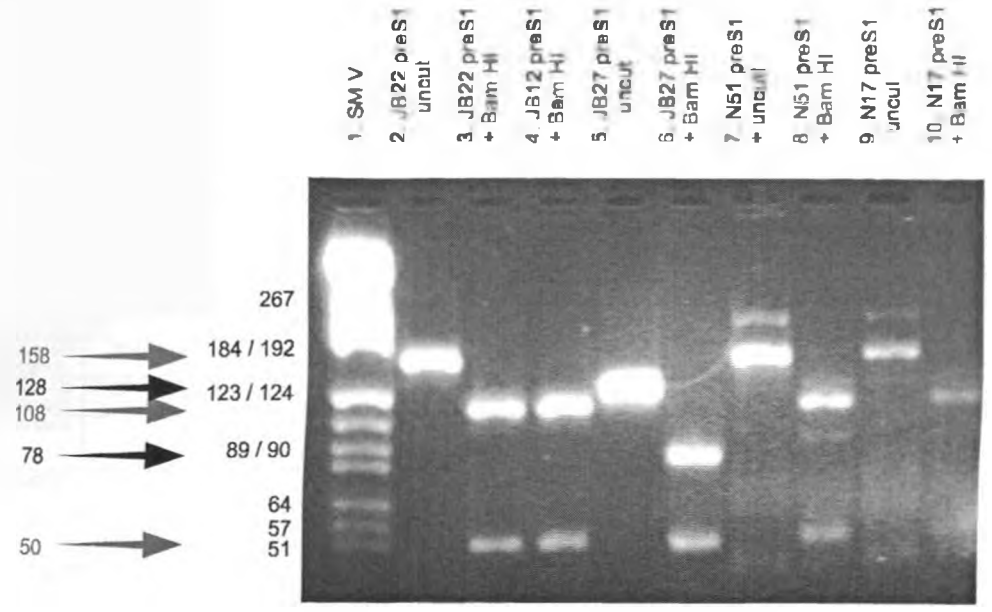


Figure 22. PS1 *BamHI* digestion. Lanes 2, 7 and 9 show the uncut 158 bp PS1 fragment while lanes 3, 4, 8 and 10 show the predicted digested fragments of 108 and 50 bp indicating that JB22, JB12, N51 and N17 are genotype E. Lane 5 shows the uncut 128 bp PS1 fragment and lane 6 shows the predicted digested fragments of 78 and 50 bp indicating that JB27 is genotype D. The larger non-specific band of about 212 bp seen in lane 7 and 9 was a common PCR artifact in the Namibian specimens

Table 10. Summary of results from South African specimens genotyped by RFLP analysis and size polymorphism, HBV genotype-specific EIA or sequence analysis

Lab ID	NIV No.. Test No.	GenBank Accession No.	Race	Sex	Age	Size	S1 BamHI (2943)	S1 TaqI (2877)	S2 BamHI (28)	S2 HpaII (75)	S2 TaqI (128)	S2 XhoI (127)	geno type (RFLP)	geno type (Seq)	geno type (EIA)
A19	92005609 W001032		B	M	Adl	161	-	-	-		+		A'		
E03	94003181 H000720			M		128	-						D		
E09	94004278 H001035		B	M		128	+		-		+		D		
E18	94005285 H001287					161	-	-	-	-	+	+	A'		
E33	94009957 H002406					161	-	-			+	-	A'		
ES1F	98024943 H006052		B	M	28	128							D		
H3414	97014914 H003414		B	M	Adl	161	-	-	-	-	+		A'		
H3516	97015594 H003516		B	F	Adl	161	-	-	-	-			A'		
H3524	H3524					161	-	-	-	-	+		A'		
O09	91025904 U000007		W	M		161			+		-	-	A		
O10	90001405 B008422		W	M	44Y	161	-		+		-	-	A	A	
S07	91004819 T003869			F		161			-	-	+	+	A'		
S08	91020704 T008627		B	F	Adl	128/ 161	-		-	-	+	+	A'+D		
P05	92017245 W004177		B	M	Adl	161	-	-	-	-			A'	A'	
A23	90007326 S003500	U87735	W	F		161	-	-	+	-	-		A	A	
A24	90006905 S002230	U87725	C	F	32	161	-		+	-	-	-	A	A	
E06	94003882 H000906	U87736	B	M		161	-	-	-	-	+	+	A'	A'	
E08	94004571 H001005	U87737	B	M	49	128	+	-	-	-	+	+	D	D	
E11	94004494 H001051	U87734	B	M		161	-	-	-	-	+	+	A'	A'	
E14	94004979 H001177	U87727	B	M		161	-	-	-	-	+	-	A'	A'	
E17	94005281 H001277	U87731	B	M	Adl	161	-	-	-	-	+	+	A'	A'	
E20	94006454 H001536	U87728	B	M		161	-	-	-	-	+	-	A'	A'	
E30	94009125 H002221	U87733	B	F		161	-	-	-	-	+	-	A'	A'	
E34	94010112 H002448	U86095	B	M		161	-	-	-	-	+	-	A'	A'	
C01	91011012 T005637	U87732	B	M	Adl	161	-	-	+	-	-	-	A	A	
C02	91019269 T008220	U88094	B	M	9	161	-	-	-	-	+	+	A'	A'	
C04	91027708 U000747	U87730	B	F		161	-	-	+	-	-	-	A	A	
P03	92013417 W003156	U87729	B	M	Adl	161	-	-	+	-	-	-	A	A	
P07	92027500 W006552	U87851	B	F	Adl	128	+	-	-	-	+	-	D	D	
P08	92030853 W007211	U87726	W	M	40Y	161	-	-	+	-	-	-	A	A	
S01	92038449 S00836	U88096	B	M		161	-	-	-	-	+	+	A'	A'	
S09	91017133 T00739C	U87738	W	F	Adl	128	+	+	-	-	+	+	D	D	
ES01	M000006		A			161	-	+	-	+			C	C	C (bks)
ES02	M000007		A			128			-				D	D	D (bksu)
ES03			A			161	-	+	-	+			C	C	C (bks)
ES04	94013588 H003197		B	M	28	161	-	-	-	-			A'	A'	E (bksu)(fg)
ES05						161	+		+				E	E	A (bsu)
ES10	96024255 H005887			M	28	128	+						D		D (bksu)
ES06	96019056 H004802		B	M											A (bsu)
ES07			A	F											C (bks)
ES11															A (bsu)
HE04															E (bksu)(fg)
HB05	97014818 H003352		B	F	Adl										A (bsu)

Table 10. Continued overleaf

Lab ID	NIV No Test No	GenBank Accession No.	Race	Sex	Age	Size	S1 Bam HI (2943)	S1 TaqI (2877)	S2 BamHI (28)	S2 HpaII (75)	S2 Taq1 (128)	S2 XhoI (127)	geno type (RFLP)	geno type (Seq)	geno type (EIA)
HB13	S7034849 H07566		A	M	Adl										B (bm)
1449		U87745	B	M										A'	
2238			B	M										D	
H510		U87740	B	M										A'	
IR		U87743	B	M										A'	
1149		U87852	C	M										A'	
7983		U87742	C	M										A'	
8017		U87744	W	M										A	
1019		U87741	W	M										A	
1389		U87739	W	M										A	
7782		U87746	A	M										A'	
WONG		U87747	A	M										B	
LEE		U87748	A	M										C	

B = Black, W = White, C = Coloured, A = Asian; Adl = Adult
empty cells indicate that these tests were not done or information was not available

absence of RE sites successfully partitioned the South African specimens. Results of the 38 urban specimens screened in this way are shown in Table 10. Sequencing and phylogenetic analysis of 25 of the 38 specimens confirmed the RFLP analysis results. Six of the 38 specimens, together with a further 6 uncharacterized specimens, were subjected to HBV genotype-specific EIA. ES04 was typed genotype A' by size polymorphism and RFLP and genotype E by EIA whereas ES05 was typed genotype E by the former method and genotype A' by the latter. Since sequence data confirmed the original results and the discrepant values for the two specimens are transposed, this discrepancy was most likely the result of an inadvertent switch or labelling error in the laboratory. The last 12 of the 56 urban specimens listed in Table 10 were genotyped by sequence analysis alone.

Thus the 56 urban Gauteng specimens from South Africa partitioned as follows: 39/56 (69.6 percent) were from genotype A of which 64.0 percent (25/39) were from subgroup A' and the remaining 14 (35.9 percent) from subgroup A. Two specimens were from genotype B and a further 2 from genotype E (2/56 or 3.6 percent each), 4 from genotype C (4/56 or 7.1 percent) and the remainder from genotype D (9/56 or 16.1 percent). Of the 57 rural samples subjected to PS1 and PS2 (nested or unnested) PCR, 64.9 percent, or 37 specimens, yielded suitable material for the digestions. The majority of these specimens (89.2 percent) genotyped as genotype A, of which only one was from subtype A, while 10.8 percent were genotype D (Vardas, 1999). The apparent absence of subtype A in the rural community would appear to indicate that subtype A' is indigenous to South Africa and that the global subtype A is a relatively new import to this country.

4.3.3 Namibian genotyping results

The HBsAg-positive specimens from the pilot study from the North East and North West of the country which gave a PS1 amplicon which was larger than expected were stored for later study but disregarded for the genotype screening study. However, three of these specimens were sent for HBV genotype-specific EIA typing and all three were typed unambiguously as genotype E by EIA. Of the 13 specimens remaining, 12 (12/13 or 92.3 percent) were from genotype E and only 1 from genotype D (1/13 or 7.7 percent). These specimens have N numbers less than 231 in Table 11.

Twenty-four of the HBsAg- and HBeAg positive specimens from the Ovambo region were successfully genotyped (JB numbers in Table 11). The 17 specimens typed by size polymorphism and RFLP analysis (9 of which were confirmed by sequence analysis and 6 by EIA) were all either genotype E (16/17 or 94.1 percent) or genotype D (1/17 or 5.9 percent). A further 7 typed by genotype-specific EIA alone were all found to be from genotype E (Table 12).

Seven of a small group of 8 HBsAg-positive chronic carriers from the Rundu region in the North East of the country (N231-N238 in Table 11) were found to be genotype E (87.5 percent) while the remaining specimen was found to be from genotype D (12.5 percent).

These results are summarized in Table 12 and give an average of 93.8 percent genotype E and 6.2 percent genotype D in the regions of northern Namibia studied.

Table 12. Summary of genotype screening results from Namibia

	Genotype E	Genotype D
Ovamboland:		
size polymorphism and RFLP analysis	16/17 (94.1 %)	1/17 (5.9 %)
HBV genotype-specific EIA	7/7 (100 %)	
Chronic carriers from NE:		
HBV genotype-specific EIA	7/8 (87.5 %)	1/8 (12.5 %)
Pilot study from NW and NE:		
size polymorphism and RFLP analysis	12/13 (92.3 %)	1/13 (7.7 %)
HBV genotype-specific EIA	3/3 (100 %)	
	45/48 (93.8 %)	3 (6.2 %)

Table 11. Summary of results from Namibian specimens genotyped by RFLP analysis and size polymorphism, HBV genotype-specific EIA or sequence analysis

Lab ID.	NIV No.. Test No	Race	Sex	Age	Size	S1 Bam HI (2943)	S1 Taq I (2877)	S2 Bam HI (28)	S2 Hpa II (75)	S2 Taq 1 (128)	S2 Xhol (127)	geno type (RFLP)	geno type (Seq)	genotype (EIA)
JB02	0720-20		F	Adl	158	+	+	+	+	+		E	E	
JB1	1109-2			child	158	+		+				E/F	E	
JB05	1161-05			child	158	+		+				E/F	E	
JB07	1127-03			child	158	+		+				E/F	E	
JB11	1246-20		F	Adl	158	+		+		+		E/F	E	
JB13	1271-07			child	158	+		+				E/F	E	
JB16	0804-02			child	158	+		+		+		E/F	E	
JB17	0808-03			child	158	+		+		+		E/F	E	
JB24	1105-01			child	158	+	+					E		
N17	UM	B	M	22	158	+	+					E		
N51	SK	B	M	22	158	+	+					E		
N52	MM	B	M	24	158	+	+					E		
N66	CH	B	M	23	158	+	+					E		
N185	FH	B	F	19	158	+	+					E		
N212	PK	B	M	19	158	+	+					E		
JB19	0998-02			child	158	+		+				E/F		
N13	AM	B	F	18	158	+						E/F		
N47	JK	B	M	23	158	+		+				E/F		
N53	EL	B	M	23	158	+		+				E/F		
N55	IN	B	F	17	158	+		+				E/F		
N59	HS	B	M	24	158		+					E/F		
N111	SH	B	M	24	158	+		+				E/F		
N118	RN	B	M	19	128	+		+				D		
JB04 JB27	0728-20 0728-03		F	Adl child	158 128	+ +	+	+ +	+			E D	E	
JB12	1265-02			child	158	+	+					E		E (bksug)
JB18 JB21	0930-05 0930-20		F	child Adl	158 158			+				E		E (bksug) E (bksug)
JB19	0998-02			child	158	+		+				E		E (bksug)
JB21 JB18	0930-20 0930-05		F	Adl child	158 158	 +		+ +				A/E/F E/F		E (bksug) E (bksug)
JB22 JB23	1081-07 1081-20		F	child Adl	158 158	+		+				E/F		E (bksug) E (bksug)
JB08 JB06	1234-02 1234-20		F	child Adl	158 158									E (bksug) E (bksug)
JB10	1242-20			Adl	158									E (bksug)
JB15 JB28	0787-20 0787-05		F	Adl child	158 158									E (bksug) D/E (bksu)
N231	97018904													E (bksug)
N232	97018905													E (bksug)
N233	97018906													E (bksug)
N234	97018907													E (bksug)
N235	97018908													E (bksu)
N236	97018909													D (bksug)
N237	97018910													E (bksug)
N238	97018911													E (bksug)
N06	WI	B	M		>158	+								E (bksug)
N07	FS	B	F		>158	+								E (bksug)
N74	PI	B	M	26	>158		-							E (bksug)

Empty cells indicate that the tests were not done or the information was not available

4.4 Discussion

At the start of this study, little genotypic information was available from sub-Saharan Africa. Using a unique combination of known RE sites, this study was able to discern not only the genotypes of HBV, but also the Cq subgroup of genotype C (Norder *et al.*, 1994) and a previously unidentified subgroup partition of genotype A, subgroup A'.

Since more than half of genotype A specimens from South Africa belonged to the newly identified subset of genotype A, subgroup A' (Bowyer *et al.*, 1997), I examined various phylogenetic trees in the literature retrospectively and found that the majority of specimens which cluster with the Philippino specimen M57663 (Estacio *et al.*, 1988) originate from Africa. The only exception, LS (Kidd-Ljunggren *et al.*, 1994), was notable in that it also originated from the Philippines. However, more recent studies from South and Central America show that 50 percent or more of the genotype A specimens from the New World also cluster into subgroup A' (Moraes *et al.*, 1996; Arauz-Ruiz, 1997a).

The results of the screening assays agreed with each other and with the sequencing results when parallel data was collected with one exception. The HBV genotype-specific EIA and the size polymorphism and RFLP assay gave transposed results for the specimens ES04 and ES05. Since sequence analysis confirmed the size polymorphism and RFLP results in both cases, I concluded that the discrepancy was probably the result of an inadvertent laboratory specimen switching error.

In the Gauteng province of South Africa I found 25.0 percent subgroup A, 44.6 percent subgroup A', 3.6 percent of genotype B, 7.1 percent genotype C, 16.1 percent genotype D and 3.6 percent genotype E. In the Eastern Cape, however, subgroup A' was even more prominent and I found 86.5 percent subtype A', 10.8 percent genotype D and only one specimen or 2.7 percent of subtype A. Based on these results, the South African distribution of serotypes reported by Courouce-Pauty *et al.* (1983) under-represents the prevalence of the *ayw* phenotype. This study found 97.7 percent of *adw* and only 2.3 percent of *ayw* whereas I have found 5 to 6 times more than this of the *ayw* genotype, genotype D, and this does not take into consideration the small proportion of serotype *ayw1* present in genotypes A and C. The sera in the Courouce-Pauty's serotype study came from black donors attending the

Natal blood transfusion centre in the Eastern Kwazulu-Natal province of South Africa (Courouce-Pauty *et al.*, 1983), and may reflect a regional difference in genotype distribution.

The only other genotype, or serotype, data available from South Africa is the control data in the molecular epidemiological single-stranded conformational polymorphism (SSCP) transmission study of Hardie *et al.* (1996; 1997) from Cape Town in the Western Province of South Africa. Phylogenetic analysis of 27 control specimens in this study revealed that 14/27 (51.8 percent) were from genotype A, of which 13 (92.9 percent) clustered with M57663 (Estacio *et al.* 1988) and were therefore from subgroup A' and only one from subgroup A (7.1 percent). Twelve of the remaining specimens were from genotype D (44.4 percent) and one from genotype C (3.7 percent). The lack of subgroup A in this control group is in accord with the findings of Vardas (1999) in the Western Cape, although the incidence of genotype D in this study is considerably higher than any figure previously reported from any region of southern Africa. On the basis of this limited data, it appears that genotype A is almost exclusively represented by subtype A' in the Cape and that genotype D is much more common in the Western Cape.

The specimens acquired from Namibia's North West and North East provinces were found to be almost exclusively from genotype E (93.8 percent) with a small percentage of genotype D (6.2 percent). No other genotype was detected. Although quite different from the distribution of genotypes found in South Africa, this result is consistent with the serotypic data of Courouce-Pauty *et al.* (1983) who found 82.3 percent of serotype *ayw4*, which is unique to genotype E, and 16.5 percent of other *ayw* serotypes in west Africa. Courouce-Pauty *et al.* (1983) also found a very small percentage (1.2 percent) of serotype *adw* in west Africa. My results are however not representative of the whole of Namibia since all of my specimens come from the northern provinces. Also 12 of the JB numbered specimens represented mother/child pairs and a large number of the RE digests were excluded from the results because their PS1 amplicon was too large. The Namibian results can therefore be regarded merely as a pilot study which gives some indication of the strains circulating in Namibia and a large scale molecular epidemiological survey of the region is in progress.

There is a growing perception that a relationship exists between genotypic differences in HBV and chronic pathogenesis in the infected liver. Li *et al.* (1993) used size polymorphism in the preS1 and core gene to partition specimens into three groups of genotypes, A, B/C and D/E and showed that genotype can influence the course of HBV disease. In particular, they showed that genotype A rarely circulates as an HBe-minus mutant, probably because a UAG nonsense mutation at codon 28 of the core gene would require a simultaneous sequence change at codon 15 from CCC to CCT in genotype A, whereas the other genotypes already have a CCT at codon 15. More recent sequence data shows that genotype F shares this predisposition with genotype A. Zhang *et al.* (1996) used Li's genotypic screen (Li *et al.*, 1993) and a hybridisation method to detect precore gene mutants using a cohort of patients on interferon therapy. Their data supports the hypothesis that precore defective HBV mutants have an increased capacity to resist exogenous alpha interferon resulting in a higher response rate in genotype A patients. Genotyping prior to interferon treatment would therefore predict antiviral response. Kao *et al.* (2000) recently reported that genotype C is associated with more severe liver disease than genotype B and may also be associated with the development of HCC in young Taiwanese.

CHAPTER 5

A UNIQUE SEGMENT OF HEPATITIS B VIRUS GENOTYPE A IDENTIFIED IN ISOLATES FROM SOUTH AFRICA

5.1 Introduction

Genotype screening indicates that 69.6 percent of urban isolates and 89.2 percent of rural isolates from South Africa partition into genotype A. Most of the remaining specimens partitioned into genotype D. The screening assay also revealed that many of the genotype A specimens from South Africa did not have the preS restriction enzyme sites predicted from GenBank sequence data. To investigate these differences I sequenced the preS2 and S gene regions of 29 isolates from patients with acute hepatitis B and from chronic carriers of the virus in the Gauteng province of South Africa and compared my data with published sequences.

Since the 9 major serotype divisions, *adw2, adw4 q , ayw1, ayw2, ayw3, ayw4, adr q , adr q* and *ayr* are based on a common *a* determinant and two mutually exclusive *d/y* and *r/w* determinants of the S gene product I also deduced the serotype of my specimens from their sequence and confirmed my predictions experimentally.

5.2 Material and Methods

5.2.1 Patients

HBV DNA was extracted from 29 South African patients, 11 with acute hepatitis B, and 18 chronic carriers of the virus. Twenty-four of the patients were male and five were female; 17 Blacks, four Coloured (i.e. Eur-African, or mixed, descent), five Whites (i.e. European, or Caucasian, descent) and 3 of Asian descent (one of Indian and two of Chinese descent). The full sequence over the preS2 and S gene was obtained from 26 of these patients, whereas two were only sequenced in the preS2 region and one in the S gene region.

5.2.2 HBV DNA extraction

HBV DNA was extracted from 200 µl serum using QIAamp rapid purification columns (Qiagen), according to the manufacturer's instructions.

5.2.3 DNA amplification

For the acute hepatitis B patients, the complete preS/S gene was amplified using sense primer 2810 (5'CACGTAGCGCCTCATTTTCTGGGTCACCATATTCT 3') and antisense primer 979 (5' CAAAAGACCCACAATTCTTTGACATACTTTCCAAT 3') to yield a 1403 bp amplicon. The proofreading enzyme *Pfu* (0.0025 U/µl) was added to the PCR reaction mix (10 mM Tris-HCl pH 9.0, 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM each dNTPs, 0.05 U/µl *Taq* polymerase [Promega], 1 µM of each primer, 0.1 percent Triton-X100) to enhance the fidelity of the polymerase reaction.

The chronic carriers were treated differently in anticipation of the need to perform nested PCR on some of these specimens. Initially the entire core and surface genes of the chronic carriers were amplified using sense primer 1738 (5'AGTTGGGGGAGGAGATTAG 3') and antisense primer 1018 (5' CCACATTGTGTAAATGGAGCAGC 3') to yield a 2485-2524 bp fragment (depending on genotype) using *TaqExpand* polymerase enzyme (Roche Diagnostics, Mannheim, Germany), which contains proofreading activity as described in paragraph 3.2.2. If no product was evident on first round PCR, a nested PCR was performed using sense primer 2833 (5'CTTGGGAACAAGAGCTACAGCAT 3') with the same antisense primer as before, Asn1018. This gave a 1396-1427 bp amplicon which included the entire preS region and all of the HBsAg gene (paragraph 3.2.2 and Table 4).

5.2.4 DNA sequencing

Oligonucleotides and excess dNTPs were removed from the PCR products either by using Magic PCR Preps (Promega) or by digestion with Exonuclease I and alkaline phosphatase according to the protocol from the USB direct sequencing kit (Amersham). The nucleotide sequence of the preS2 (from nucleotide -11 to 154) and/or the surface gene which codes for the middle or major surface antigen (from nucleotide 155 to 845) of all amplicons was determined by the dideoxy chain termination method (USB Sequenase direct sequencing kit, Amersham) in both forward and reverse directions.

5.2.5 Serotyping

Serotyping was either done using the HBsAg subtype kit and method from the Institute of Immunology Co., Ltd., Tokyo, Japan or deduced from relevant amino acids (predicted from the nucleotide sequence) in the HBsAg gene as described in the literature (Okamoto *et. al.*, 1987; 1989; Norder *et al.*, 1992).

5.2.6 Data analysis

Twenty-six South African isolates were compared with 43 sequences from GenBank, complete over the relevant 846 nucleotides. The remaining three South African isolates, lacking some sequence information at either end of the fragment were analyzed together with 23 "preS2 only" or 15 "major surface gene only" sequences from GenBank, which, together with my 26 full sequences and the 43 GenBank sequences, enabled a comparison of 93 specimens in the preS2 region and 86 in the major surface gene region. All sequences used in this study are referenced in Table 13. Mutational "hotspots", or regions of variation, within the sequenced region were determined by storing all sequence information in a database (dBase III+, Ashton Tate) and then calculating the frequency of each base at each nucleotide position using a customized programme (computer programmes and results discussed in Chapter 6). Dendrograms were generated using the PHYLIP (Phylogeny Inference Package) version 3.5c DNAPARS, FITCH, DNAML, DRAWTREE and DRAWGRAM software programs (Felsenstein, 1993).

5.3 Results

Using mutational "hotspot" data, the degree of diversity of important preS2 and S domains was calculated, both at the nucleotide and amino acid level (Table 14).

Fourteen of the 20 South African specimens serotyped using monoclonal antibodies directed against the *a, d, y, w* and *r* determinants of the surface antigen were found to be serotype *adw*, four were found to be serotype *ayw*, one serotype *adr* and one, A40, had the mixed serotype *adyw*. The serotype of the 20 experimentally serotyped South African specimens are marked with an asterisk in Table 13. Where the text specifically states the serotype of the reference specimens they are similarly marked with an asterisk. The serotype of the other 9 sequenced South African specimens, and all reference specimens for which the

Table 13. Designation of 110 HBV isolates used to characterize the subgroup A' of genotype A

No.	Patient / strain	Accession No. / reference	Genotype	Serotype	Origin
A01	HHVBAMAM	X75669	A	*ayw1**	Cameroon
A02	HPBS2ADW	M12346	A	*adw**	The East
A03	S50225	S50225	A	*adw*	UK
A04	HPBSSWT	M74498	A	*adw2*	FranceA
A05	HPBSSB1MUT	M74499	A	*adw2*	France
A06	HHVBA	X75666	A	*adw2**	France
A07	HVHEPB	X51970	A	*adw**	Germany
A08	HBVXCPS	X70185	A	*adw2**	Germany
A09	HPBS	M21030	A	*adw2**	Mozambique
A10	LS	Kidd-Ljunggren†	A	*adw2**	The Philippines
A11	JOA	Kidd-Ljunggren†	A	*adw2**	Poland
A12	MP	Kidd-Ljunggren†	A	*adw2**	Poland
A13	HUMPRECX	L13994	A	*adw**	Eastern Europe
A14	1385	U87739‡•	A	*adw**	South Africa
A15	A24	U87725‡	A	*adw*	South Africa
A16	P08	U87726‡	A	*adw*	South Africa
A17	E14	U87727‡	A	*adw*	South Africa
A18	H510	U87740‡•	A	*adw**	South Africa
A19	E20	U87728‡	A	*adw**	South Africa
A20	7983	U87742‡•	A	*adw**	South Africa
A21	P03	U87729‡	A	*adw*	South Africa
A22	O04	U87730‡	A	*adw**	South Africa
A23	E17	U87731‡	A	*adw**	South Africa
A24	O01	U87732‡	A	*adw*	South Africa
A25	E30	U87733‡	A	*adw**	South Africa
A26	1019	U87741‡•	A	*adw**	South Africa
A27	1149	U87852‡•	A	*adw**	South Africa
A28	8017	U87744‡•	A	*adw**	South Africa
A29	1449	U87745‡•	A	*adw*	South Africa
A30	IR	U87743‡•	A	*adw**	South Africa
A31	O02	U88094‡	A	*adw**	South Africa
A32	E11	U87734‡	A	*adw*	South Africa
A34	E34	U88095‡	A	*adw**	South Africa
A36	A23	U87735‡	A	*adw*	South Africa
A38	S01	U88096‡	A	*adw*	South Africa
A39	E06	U87736‡	A	*ayw **	South Africa
A40	7782	U87746‡•	A	*adyw **	South Africa
A41	HPBVSAG2	M54898	A	*adw**	Taiwan
A42	HBADWZCG	M57663	A	*adw2**	The Philippines
A43	HBVADW	V00866	A	*adw2**	USA
A44	HBVADW2	X02763	A	*adw2**	USA
A45	HBVPRESS	X69458	A	*ayw **	Zimbabwe
A46	HPBSSB1REC	M74500	A	*adw2*	France
B01	HPBADW3/ pIDW420	D00331	B	*adw3**	Indonesia
B02	HPBADW1/ pJDW233	D00329	B	*adw2**	Japan
B03	HPBADWZ/ pMND122	M54923	B	*adw2**	Indonesia
B04	HPBADW2/ pODW282	D00330	B	*adw2**	Japan
B05	WONG	U87747‡•	B	*adw**	South Africa
B06	ISWARI	Lauder§	B	*adw*	
B07	HVBBS	X75660	B	*ayw1**	
C01	OKAADW	Lauder§	C	*adw**	
C02	HPBADR1CG/ pADR1	M38454	C	*adr**	China
C03	CP	Kidd-Ljunggren†	C	*adr**	China
C04	S62754	S62754	C	*adw**	The East
C05	HPBSAYRA	M17688	C	*ayr**	The East
C06	HHVBCS	X75792	C	*adr**	France
C07	DD	Kidd-Ljunggren†	C	*adr**	Indonesia

Table 13 continued overleaf

No.	Patient / strain	Accession No. /	Genotype	Serotype	Origin
C08	HPBCGADR	D12980	C	*adr**	Japan
C09	HBVADR/ ONOADR	V00867	C	*adr**	Japan
C10	HPBETNC	L08805	C	*adr**	Japan
C11	HPBADRA	M12906	C	*adr**	Japan
C12	HBVADR4	X01587	C	*adr**	Japan
C13	HPBHBSAGA	M23805	C	*adr*	Japan
C14	HPBHBSAGB	M23806	C	*adr**	Japan
C15	HPBHBSAGD	M23808	C	*adr**	Japan
C16	HEHBVAYR	X04615	C	*ayr**	The East
C17	HPBCGADR	M38636	C	*adr**	Korea
C18	HBVADRM	X14193	C	*adr**	Korea
C19	KRJ	Kidd-Ljunggren†	C	*adr**	Korea
C20	X75665	X75665	C	*adrq-**	New Caledonia
C21	491	Kidd-Ljunggren†	C	*adr**	New Zealand
C22	651	Kidd-Ljunggren†	C	*adr**	New Zealand
C23	29	Kidd-Ljunggren†	C	*adr**	New Zealand
C24	HHVCCHA	X75656	C	*adrq-**	French Polynesia
C25	LEE	U87748‡•	C	*adr**	South Africa
C26	AP	Kidd-Ljunggren†	C	*adr**	Thailand
D01	HPBAYW pHB320	J02203 V01460	D	*ayw3**	France
D02	HHVBDS	X75668	D	*ayw3**	France
D03	HHVBDS	X75662	D	*ayw2**	France
D04	HBVORFS/ C1005	X72702	D	*ayw**	Germany
D05	PC	Kidd-Ljunggren†	D	*ayw**	India
D06	NB	Kidd-Ljunggren†	D	*ayw**	Iran
D07	HBVAYWCI/ patient CI	X65258	D	*ayw**	Italy
D08	HBVAYWE/ patient E	X65259	D	*ayw**	Italy
D09	HBVAYWC/ patient C	X65257	D	*ayw**	Italy
D10	HBVAYWMCG/ patient MG	X59795	D	*ayw**	Italy
D11	HBVDNA	X68292	D	*ayw**	Italy
D12	TY5	Kidd-Ljunggren†	D	*ayw2**	Lebanon
D13	XXHEPAV	X02496	D	*ayw**	Northern Europe
D14	376	Kidd-Ljunggren†	D	*ayw**	New Zealand
D15	567	Kidd-Ljunggren†	D	*ayw2**	New Zealand
D16	FEMI	Kidd-Ljunggren†	D	*ayw**	Romania
D17	HPBHBSAG	M12393	D	*ayw**	Eastern Europe
D18	HPBADYW	J02202	D	*adyw* *	UK
D19	MAJ	Kidd-Ljunggren†	D	*ayw2**	Somalia
D20	E08	U87737‡	D	*ayw**	South Africa
D21	S09	U87738‡	D	*ayw**	South Africa
D22	P67	U87851‡	D	*ayw**	South Africa
D25	HPBHBVAA/ HBV alpha1	M32138	D	*ayw2**	Turkey
D26	AN	Kidd-Ljunggren†	D	*ayw**	Turkey
E01	YA	Kidd-Ljunggren†	E	*ayw4**	Liberia
E02	JAM	Kidd-Ljunggren†	E	*ayw4**	Liberia
E03	HPBVAR	L24071	E	*ayw**	Northern Europe
E04	HHVBE4	X75664	E	*ayw4**	Senegal
E05	HPBVCG	D00220	E	*adw2**	West Africa
E06	HHVBBAS	X75657	E	*ayw4**	West Africa
F01	HBVADW4A	X69798	F	*adw4**	Brazil
F02	HHVBF	X75663	F	*adw4q-*	Colombia
F03	HHVBFFOU	X75658	F	*adw4q-*	France
F04	HHVBFS	X75661	F	*adw4q-*	France

* South African isolates that were serologically subtyped or specimens whose serotype was specifically stated in the reference; all other serotypes were deduced from the sequence

† Sequences from Kidd-Ljunggren *et al.* (1994)

‡ Sequences taken from Bowyer *et al.* (1997) • indicates that the sequencing was done by co-author van Staden

§ Sequences from Lauder *et al* (1993)

Table 14. Percentage diversity at the nucleotide and amino acid level across the different domains spanning the preS2/S gene open reading frame

Nt no.	Nucleotide diversity (%)	PreS2/S aa spanned	Amino acid diversity (%)	Domains of preS2 and S
3211-154	36.3 %	1 - 55	43.6 %	preS2 gene
155-175	25.0 %	1 - 7	42.8 %	Amino terminal of S
176-238	11.0 %	8 - 28	29.0 %	First membrane-spanning helix of S
239-391	15.3 %	29 - 79	23.0 %	Intracellular loop of S
392-448	6.0 %	80 - 98	4.5 %	Second membrane-spanning helix of S
449-631	20.3 %	99 - 159	20.0 %	Hydrophillic extracellular loop
632-706	13.3 %	160 - 184	28.0 %	Third membrane-spanning helix of S
707-718	0.00 %	185 - 188	0.00 %	Short region between helices
719-781	19.0 %	189 - 209	33.3 %	Fourth membrane spanning helix of S
782-835	22.2 %	210 - 226	29.0 %	Carboxy tail

serotype was not specifically stated, was deduced from the predicted amino acid sequence and these are listed without an asterisk in Table 13. The additional 9 South Africans were all found to be serotype *adw*.

The 68 complete preS2/S sequences were analyzed using DNAPARS and DRAWGRAM to produce a horizontal phylogram (results not shown). The same sequences were then used as input for DNADIST and the output distance matrix fed into FITCH and DRAWTREE to produce the tree shown in Figure 23. The South African samples cluster predominantly with known genotypic groups A and D. Of the five serotype *ayw* South African isolates, three clustered into genotype D and two (including the specimen with mixed serotype) into genotype A. The rest of the specimens clustering with genotype A were serotype *adw*. Sequence data from two patients, both of Chinese origin, did not cluster with the rest of the specimens but clustered into genotype B (*adw*) and genotype C (*adr*), respectively.

To improve the accuracy of the trees, the two genotypic groups of interest, namely, genotype A and genotype D were analyzed individually using the non-distance matrix program DNAML (Figure 24). Both Figure 23 and Figure 24(a) show clearly that genotype A splits into two distinct subgroups. Forty-three percent of southern African genotype A isolates (A14, A15, A16, A21, A22, A24, A26, A28 and A36) cluster with published sequences forming the first subgroup of A, while the remaining 57 percent (A17, A18, A19, A20, A23, A25, A27, A29, A30, A32, A39, A40) cluster separately to form a discrete

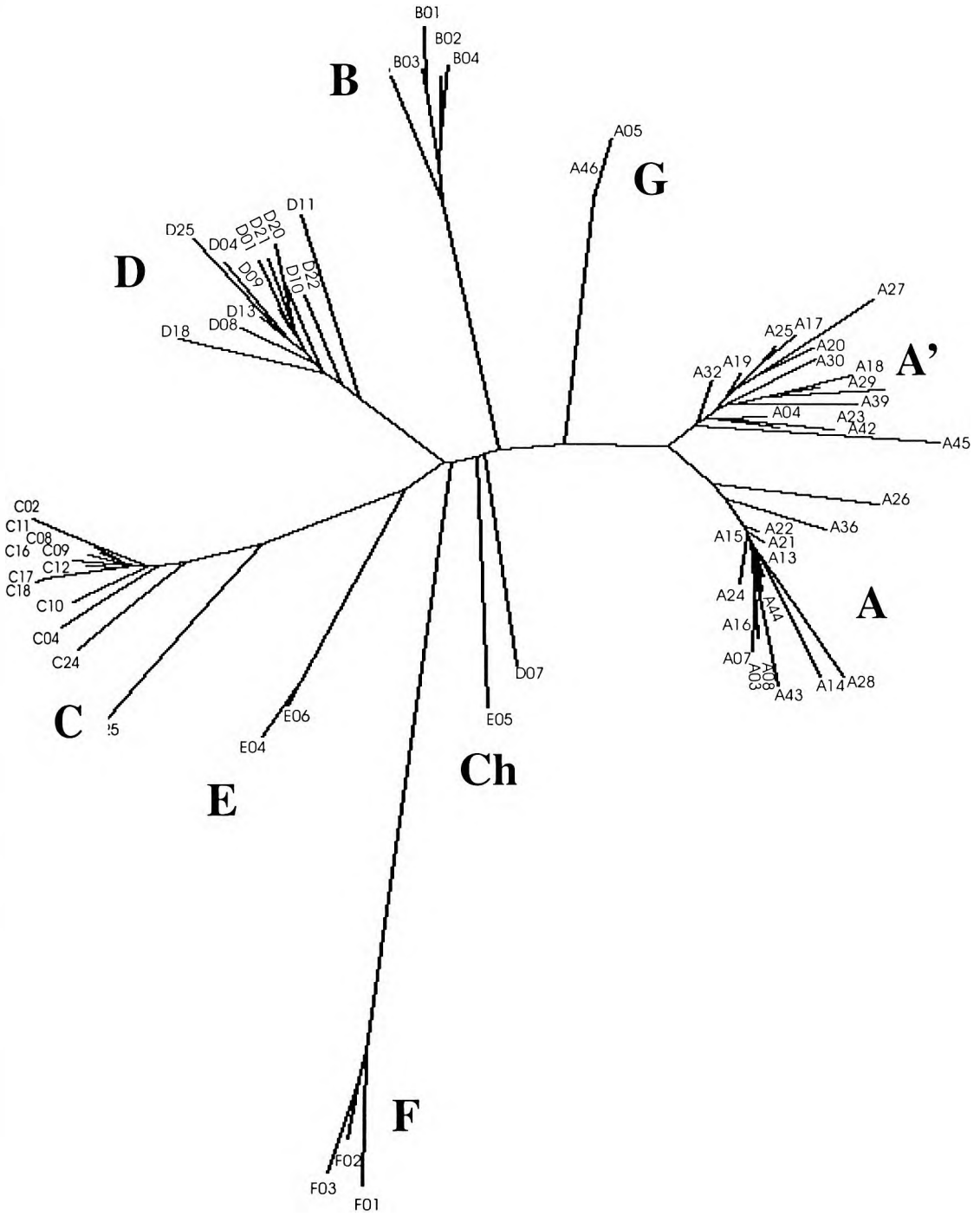


Figure 23. Phylogenetic distance tree produced using the PHYLIP programmes FITCH and DRAWTREE from preS2/S sequence data from 26 South African sequences and 42 global sequences from GenBank

genotypic segment which I referred to as subgroup A'. No such polarisation occurs in the DNAML tree of genotype D specimens (Figure 24b). Also clustering in A' were three isolates from GenBank, A45 (a Zimbabwean isolate), A04 (a French isolate; Tran *et al.*, 1991) and A42 (a clone isolated from the Philippines; Estacio *et al.*, 1988). A46 and A05, which are both variants of A04, are distant from the rest of subgroup A'. They were both grouped recently with the new putative genotype of HBV, genotype G (Stuyver *et al.*, 2000).

When the preS2 sequences from Kidd-Ljunggren *et al.* (1994) and the S only sequences from Norder *et al.* (1992) are included in the analyses the new subgroup expanded from 15 to 21 (including 62.5 percent of the southern African genotype A specimens). Seventeen of these (81 percent) are African in origin since the Mozambiquan PLC/PRF/5 hepatoma cell line sequence as well as a sequence from Cameroon and the three incomplete South African sequences cluster with this group. Of interest, the only other addition to the group is Kidd-Ljunggren's specimen LS (A10) from the Philippines (Kidd-Ljunggren *et al.*, 1994).

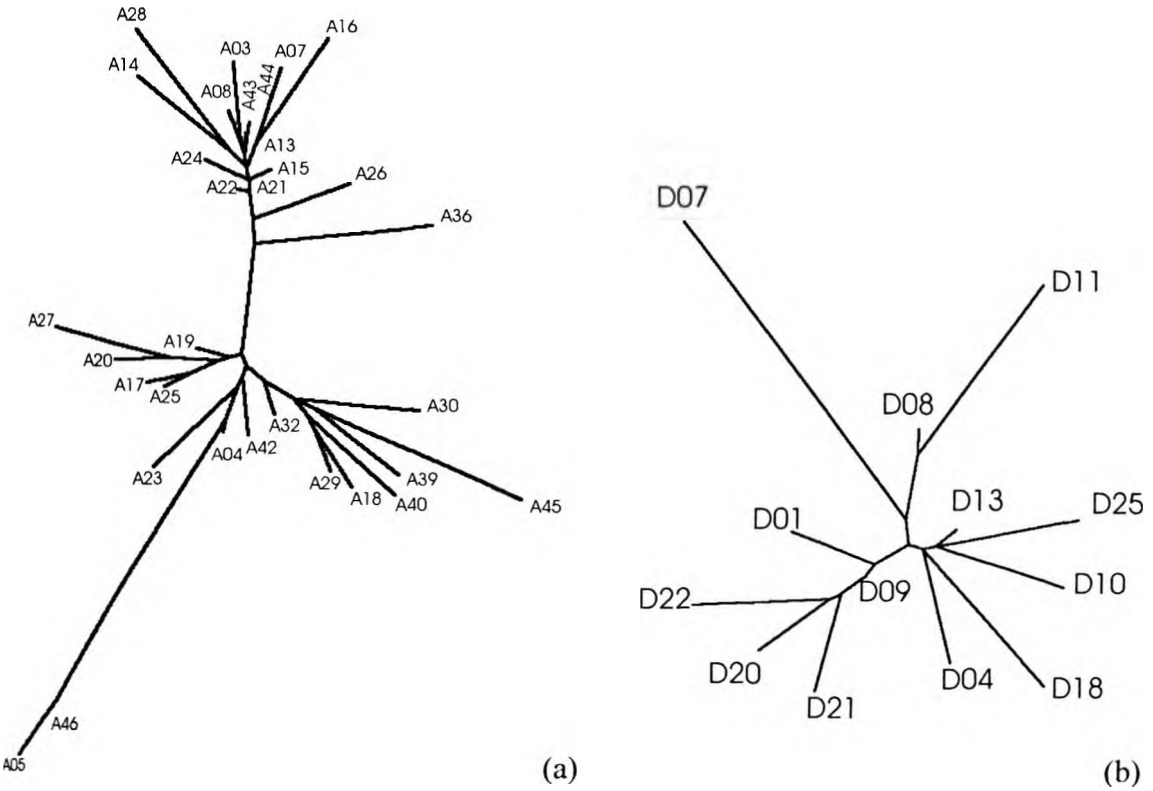


Figure 24. Phylogenetic trees produced by the PHYLIP non-distance matrix program DNAML using preS2/S sequence data. (a) 21 South African and 10 GenBank genotype A sequences and (b) three South African and 10 GenBank genotype D sequences

Average pair-wise intra- and intergroup nucleotide and amino acid divergence was calculated for 32 complete preS2/S gene regions. From this summary information [Figure 25 (a)], it can be seen that in this region genotype A is more similar to genotypes D (6.6 percent) and E (6.8 percent) than it is to the Eastern genotypes B (7.4 percent) and C (8.1 percent), and most different from genotype F (11.0 percent). Pair-wise analysis of nucleotide divergence of genotype A sequences of HBV DNA showed a mean difference between these two subgroups of genotype A of 3.8 percent whilst the difference within each subgroup was 1.7 percent (subgroup A) and 2.9 percent [subgroup A'; Figure 25(b)].

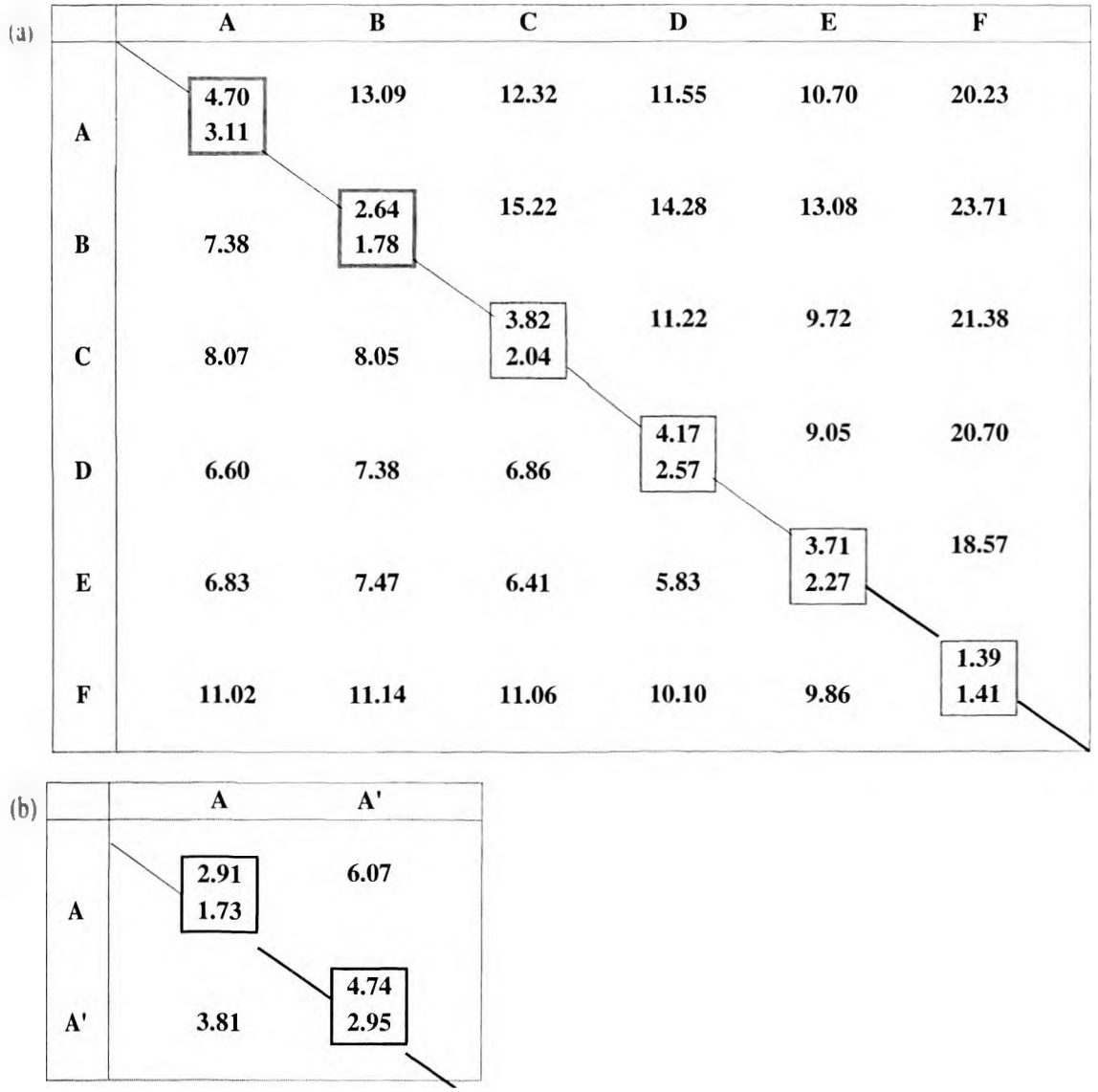


Figure 25. Mean intra– and intergroup percentage divergence in nucleotide and amino acid sequences of the preS2/S gene region of (a) the 6 genotypic groups A to F and (b) the two subgroups of A. The intragroup percentage difference is boxed. The nucleotide divergence is shown on the left, and the amino acid divergence to the right of the diagonal line

Figure 26 shows the sequence alignment of the preS2 and/or S region of HBV from the 29 South African specimens together with 80 sequences from GenBank or the literature. Only variable amino acid sites are shown in the Figure where variable amino acid sites are defined as positions at which the predominant amino acid was present at a frequency of less than 95 percent. From Figure 26, it can be seen that sites unique to genotype B only occur at amino acid 36 and 46 in the preS2 region and amino acid 5, 56, 57, 64, 85 and 213 in the S region. Genotype F-specific sites occur at amino acid 32, 42, 51 and 54 in the preS2 region and 18, 19, 45, 61, 158, 178, 183, 206, 220 and 225 in the S region. All genotype F specific sites are either in the amino-terminal or the fourth helix region of the major surface antigen. Genotype C determinants occur at surface amino acid 4, 126 and 160. Residue 45 of HBsAg, ^{45}Ser, is unique to Genotype A. Genotype D is defined in the surface gene by ^{45}Thr, ^{46}Thr and ^{134}Tyr. Genotype E did not have any unique definitive sites but shared three sites with genotype F, preS2 site 41 and surface antigen sites 127 and 140. A46 and A05 show a series of common mutations setting them apart from the rest of group A' at preS2 sites 31, 37, and 49 and surface antigen sites 45, 49, 68, 114, 134 and 143. PreS2 ^{31}I and HBsAg ^{45}V are unique to these two specimens which have recently been identified as members of the newest genotype of HBV, genotype G (Stuyver *et al.*, 2000).

5.4 Discussion

Using the preS2/S gene nucleotides, which comprise 26.3 percent of the total genome, I have genotyped 26 HBV isolates from acutely and chronically infected individuals. An additional three PCR amplicons lacking either the preS2 region (one specimen) or part of the S gene (two specimens) were also successfully genotyped. The degree of diversity (Table 14) of the preS2 region, at both the nucleotide and amino acid levels, is greater than that of the S region. When the preS2 region alone (constituting 5.1 percent of the genome) was used for the analysis, this short stretch of 165 nucleotides was found to be sufficient for the definition of the HBV genotypic groups.

Trees drawn using the 681 nucleotides of the S gene or a combination of both the preS2 and S gene were identical to the tree using preS2 nucleotide sequence alone with one exception. Specimen, D07, the GenBank HPBAYWCI mutant (Lai *et al.*, 1991), which was isolated from a patient with HBsAg-negative chronic liver disease, groups with genotype D in the

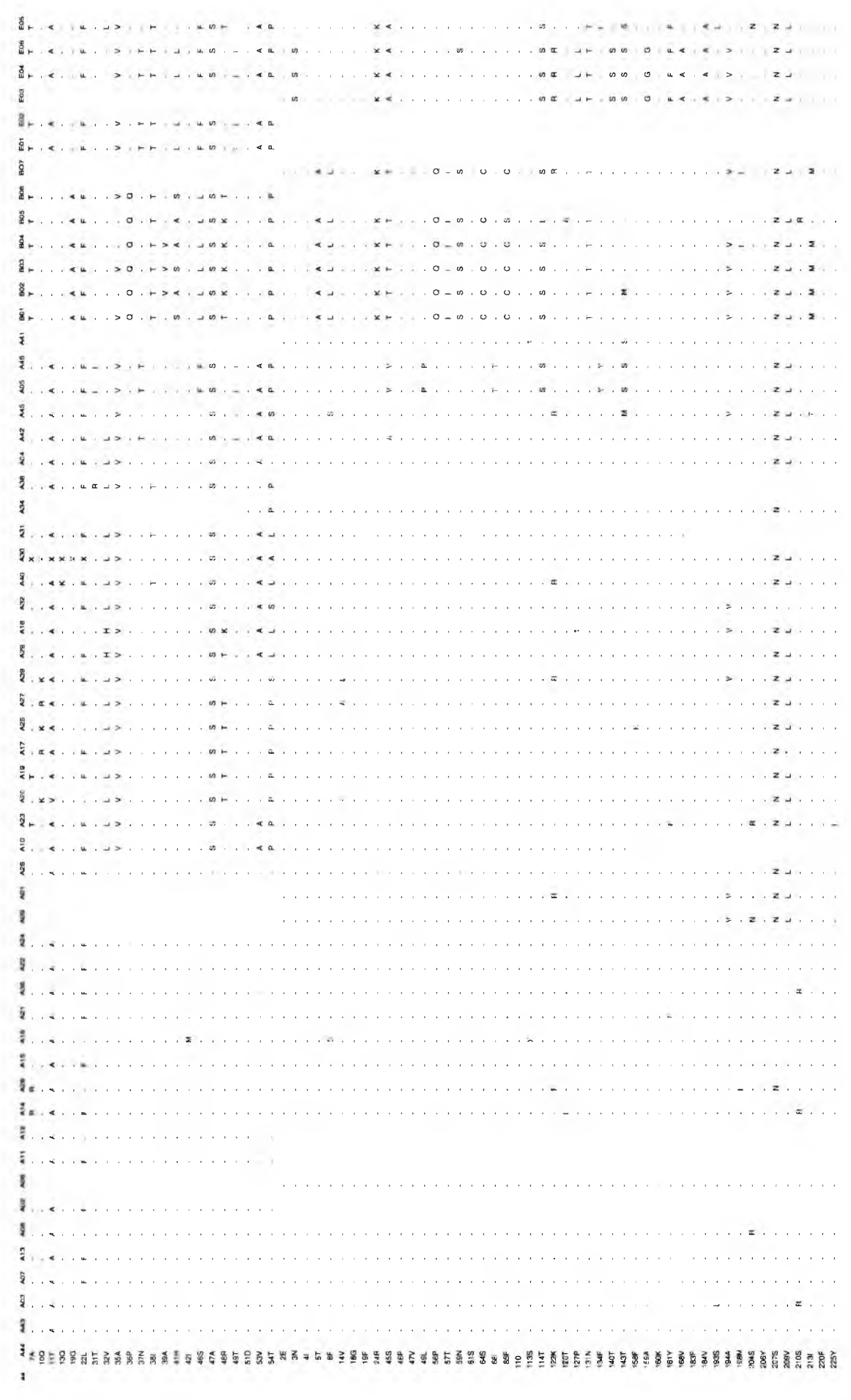


Figure 26. continued overleaf

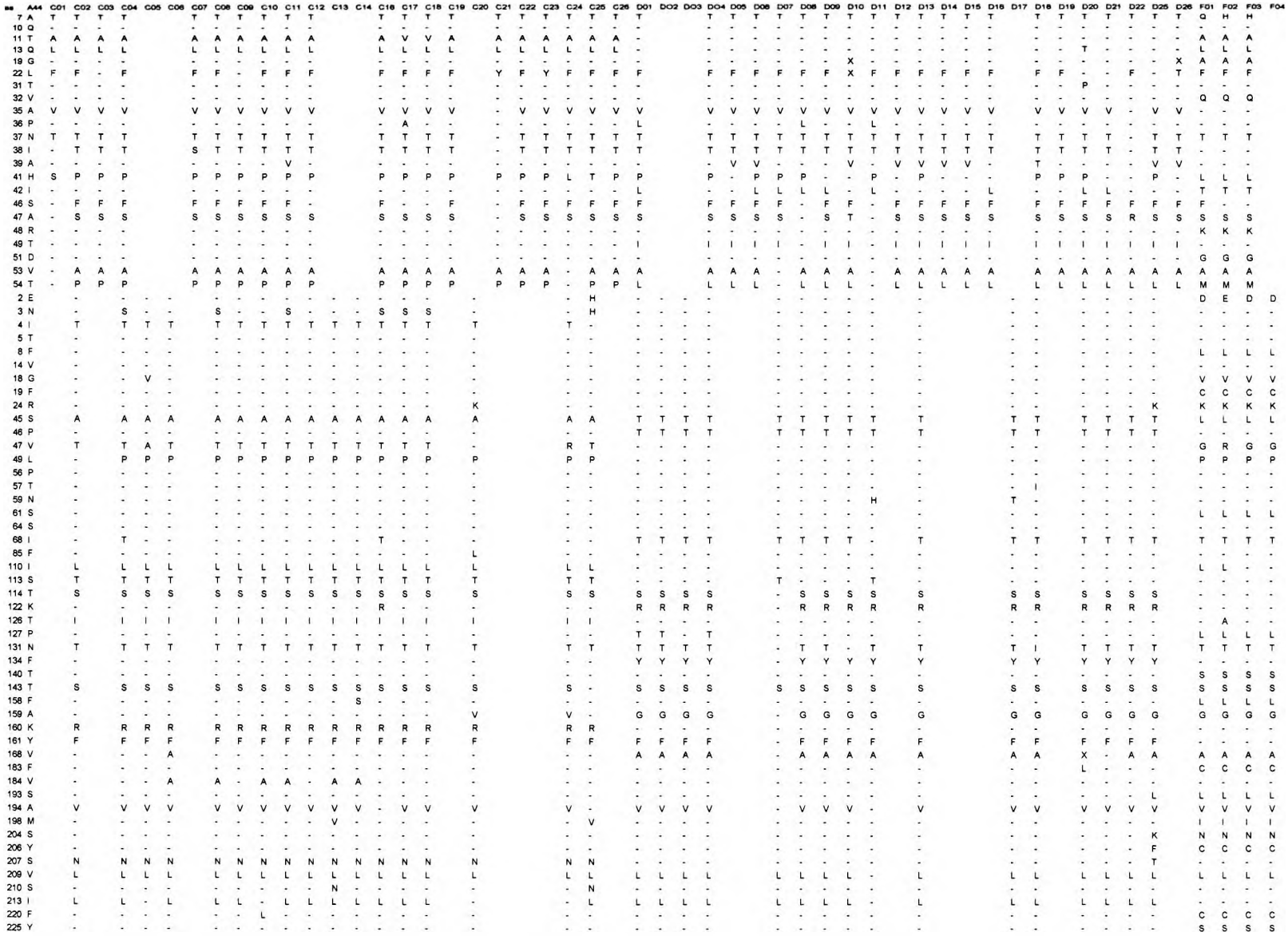


Figure 26. PreS2/S amino acids present at HBV variant sites in representative specimens from genotypes A to F. The amino acids are numbered in ascending order from amino acid 1 to 55 in the preS2 region and 1 to 226 in the S region

preS2 tree and genotype A in the surface antigen tree. It thus behaves like a putative recombinant where a recombinant is defined as a specimen which falls into "significantly discordant positions in different trees" (Hahn *et al.*, 1995). When the full preS2/S sequence was used in the analysis, it becomes an outlier, separated from known genotypic groups (Figure 23).

In the past, specimens were categorized by serotype but, since serotype specificities are defined by point mutations, it is far more informative to sequence longer stretches of the genome and then assign the isolate to a genotype by phylogenetic analysis. However, since my sequenced region included the part of S containing the *d/y* and *w/r* determinants I serotyped my specimens as well. A lysine (Lys) to arginine (Arg) switch at amino acids 122 and 160 of the S gene defines the *d/y* and *w/r* serotype specificities, respectively (Okamoto *et al.*, 1987; 1989). Isoleucine 126 (^{126}Ile) is thought to abolish the *w* determinant and as such defines the *r* determinant. No other sites previously regarded as *d/y* or *w/r* determinants have consistently maintained their specificity as global sequence data accumulate. Residues 127 and 134, however, did consistently discriminate between the *w* determinants in the specimens examined. As previously described, proline 127 (^{127}Pro) indicates *w1* or *w2*, threonine (^{127}Thr) signifies *w3* and, since both genotypes E and F have leucine (Leu) at this position, without ambiguity, ^{127}Leu most probably determines *w4*. *w2* can be discriminated from *w1* by examining amino acid 134 where phenylalanine (^{134}Phe) is indicative of the former and tyrosine (^{134}Tyr) of the latter. Based on these sites, all my genotype A *adw* specimens were of specificity *w2*.

The *ay* specificity was detected in 6 of the 30 southern African specimens in this study. These included the three South African genotype D specimens (namely, D20, D21 and D22) as well as the two genotype A specimens, A39 (*ayw*) and A40 (*adyw*), and the Zimbabwean specimen, A45. This 20 percent occurrence of the *y* specificity is 8.6 times higher than the previously reported figure of 2.3 percent from serotype analysis of specimens from the Kwazulu-Natal province of South Africa (Courouce-Pauty *et al.*, 1983).

Amino acid alignment revealed sites common to variant genotype A sequences, some of these sites have not been reported previously. To some extent variations specific to a single genotypic group tended to cluster together in the same region of the genome. ^{204}Asn

previously defined as a genotype F determinant (Norder *et al.*, 1994) was present in two non-F isolates, A09 and E05. Two of the three genotype C determinants, amino acids 126 and 160, define the *r* determinant in the extracellular loop.

^{22}Leu shows considerable variation and forms part of the immunodominant region extending from preS2 amino acid 12 to 24 reported to be involved in the primary antibody response (Meisel *et al.*, 1994). ^{11}Thr (in genotype B and D) and ^{13}Leu (in genotype C and F) occur in a region of the preS2 gene, amino acid one to 15, which is both HLA class I and class II restricted (Chisari *et al.*, 1995) and the variable residues 7 and 22 are at T-cell receptor contact sites.

Of primary interest to me was the series of variations in the South African genotype A sequences which set subgroup A' apart from the rest of genotype A, which was otherwise well conserved. These tended to cluster in the preS2 region of the genome. Also grouping with the unique South African specimens was the Zimbabwean isolate (A45), the French isolate (A04), and clone pFDW294 (A42) from the Philippines, previously observed to be most similar to the Zimbabwean isolate. When the preS2 sequences from Kidd-Ljunggren *et al.* (1994) and the S only sequences from Norder *et al.* (1992) are included in the analyses 81 percent of the new subgroup are specimens of African origin. Of interest, the only other addition to the group is Kidd-Ljunggren's specimen LS (A10) also from the Philippines (Kidd-Ljunggren *et al.*, 1994).

In the preS2 region, five of the South African specimens show a change at the TCR contact site at amino acid 7. In two ^{7}Ala is changed to ^{7}Arg and in three others it is changed to ^{7}Thr This is not, however, a subgroup-specific change as the former pair of specimens are not part of the new subgroup. A ^{7}Thr was also used before as a determinant for genotype D. At amino acid 10, glutamine (Gln) is changed to Lys or Arg in five of the specimens and the amino acids 7 to 22 are deleted in another. The change at amino acid 10 subdivides the subgroup A' of genotype A as the five specimens with this change (A19, A27, A20, A17 and A25, see Figure 26) also share specific changes at amino acids 48, 53 and 54. PreS2 amino acids 32, 35, 47, and 54 and surface antigen residues 207 and 209 split segment A' from the rest of genotype A. A01 (Cameroon) and A09 (Mozambique) have these latter two A' determinant sites. Amino acids 47 (Ala to Ser) and 48 (Arg to Thr/Lys) which define

group A' were part of the genotype-dependent recognition site at the carboxy terminus of the preS2 sequence, amino acid 42 to 49 (Meisel *et al.*, 1994).

Cysteines at sites vital for the maintenance of the antigenicity of the major surface antigen (codons 48, 65, 69, 121, 124, 137, 138, 139, 147 and 149; Antoni *et al.*, 1994) were well conserved in my samples and no changes at other previously reported vaccine escape mutant sites were found in this database (methionine 133 [^{133}Met], Oon *et al.*, 1995; ^{141}Lys, Whittle *et al.*, 1991; ^{142}Pro, Ashton-Rickardt and Murray, 1989; alanine 144 [^{144}Ala], Harrison *et al.*, 1991; ^{145}Gly, Waters *et al.*, 1992; Fujii *et al.*, 1992).

In the phylogenetic trees (Figures 23 and 24) some specimens such as the mutant, A05, and recombinant, A46, referred to by Tran *et al.* (1991) as well as the Zimbabwean variant, A45, show considerable variation from the rest of the specimens. A46 and A05 show a series of common mutations setting them apart from the rest of the specimens and these two specimens have recently been characterized as members of the newest genotype of HBV, genotype G (Stuyver *et al.*, 2000).

Analysis of the local sequence data clarified the restriction enzyme site discrepancies which I had noticed in my screening assay, and enabled me to modify the protocol to include identification of this subgroup of genotype A into which the majority of local specimens cluster. By gaining an understanding of highly conserved sites (e.g. amino acids 86 to 98, a region containing two HLA-A2 epitopes), local variation and highly variant regions of the HBV genome, I am better able to design PCR and sequencing protocols for our local situation. As viral and patient data from all corners of the world accumulates I hope to better understand how the virus and host interact and the reasons behind the different disease manifestations observed in different individuals following HBV infection.

CHAPTER 6

MAPPING THE "HOTSPOTS", THE DEFINITIVE SITES AND THE DISTRIBUTION OF INDIVIDUAL VARIATION ON THE HBV GENOME

6.1 Introduction

Identification of highly conserved regions of the HBV genome facilitates the design of primers for HBV PCR and sequencing. More importantly, nucleotide sequence variation affects the encoding of amino acids, their expression as HBV antigens and their interaction with the host's immune system. While it is interesting to evaluate the variation that a specific amino acid can undergo in carrying out a biologic function, failure to understand the variations in the strains circulating in a geographic region can lead to failure of diagnostic tests and, in extreme cases, vaccine escape. This study was designed to categorize the variation which defines the 6 genotypes of HBV at both the nucleotide and amino acid level. Definitive change is defined as change which is present in the majority of the members of a genotype or subgroup. Any one site of definitive change is not always unique to a particular genotype or subtype but the sum of the changes does define each subgroup uniquely. In addition to genotype-specific change, I also examined individual variation, non-definitive or random change, between specimens. Domingo (1998) defined a quasi-species as a dynamic distribution of non-identical but closely related mutant and recombinant viral genomes. Similarly change within an individual or between individuals of the same genotype can be either random, in which the heterogeneity at one site is unrelated to that at subsequent sites, or occur as mosaic fragments which have consecutive sites definitive of a different genotype to that identified for the specimen when other regions of the genome are examined. The latter is an indication that the heterogeneity was introduced by a recombination event. Finally, I examined the degree to which published linear epitopes and their anchor residues were conserved within the predicted HBV proteins and comment on the possible effect of HBV heterogeneity on the host immune response.

6.2 Materials and Methods

6.2.1 Patients

Sixty-five HBV full-genome nucleotide sequences, representing each of the genotypes A to F, were obtained from GenBank. For simplicity, specimens were coded according to their position in the bootstrap re-sampling tree. These codes are used throughout the text and are shown below together with their corresponding GenBank accession numbers (bold type): A01 **X51970**; A02 **E00010**; A03 **X02763**; A04 **X70185**; A05 **Z72478**; A06 **Z35717**; A07 **J02201**; A08 **L13994**; A09 **S50225**; A'01 (see **M74498** for partial sequence and Kremsdorf *et al.*, 1996 for full sequence); A'02 **M57663**; D01 **Z35716**; D02 **X72702**; D03 **X97848**; D04 **X97849**; D05 **X80925**; D06 **X80926**; D07 **X80924**; D08 **M32138**; D09 **X59795**; D10 **L27106**; D11 **X02496**; D12 **Y07587**; Dm13 **J02203**; Dm14 **U95551**; Dm15 **X65257**; Dm16 **X65258**; Dm17 **X65259**; Dm18 **X68292**; E01 **X75664**; E02 **X75657**; F01 **X69798**; F02 **X75663**; F03 **X75658**; E03 (also the ROOT) **D00220**; B'01 **X98077**; B'02 **D00330**; B'03 **X97851**; B'04 **X97850**; B05 **D50522**; B06 **D50521**; B07 **D23677**; B08 **D23678**; B09 **D23679**; B10 **D00329**; B"11 **D00331**; B"12 **M54923**; Cq⁻01 **X75656**; Cq⁻02 **X75665**; C03 **D00630**; C04 **D23684**; C05 **D23681**; C06 **D23680**; C07 **X01587**; C08 **M38454**; C09 **V00867**; C10 **X04615**; C11 **M12906**; C12 **D12980**; C13 **D16667**; C14 **D16666**; C15 **D50489**; C16 **D23682**; C17 **D23683**; C18 **L08805**.

The specimens came from 59 patients with acute, chronic or fulminant disease outcomes, three infected chimpanzees and one cell line (Dm14). Two of the chimpanzees, A07 and C18, had been recently infected with interesting patient isolates in the laboratory while the third, E03, represented long-term, possibly perinatal, natural infection. Two patients are represented by two specimens taken at different stages of disease pathogenesis (C05=C06 and C16=C17).

Subgroups were defined by an intra-subgroup divergence of four percent or less and an inter-subgroup divergence of 4 to 8 percent. Thus clusters were combined into higher order clades whose component isolates fitted this definition. Retaining the standard cut-off of 8 percent divergence to define genotypes, the 65 specimens were classified into a series of 11 subtypes within the 6 genotypes. Genotype A was divided into the subgroups A and A'; Genotype B was divided into the three subgroups B', B and B"; Genotype C was divided into the two subgroups C and Cq⁻; Genotype D was divided into a core subgroup, D, and the outliers of this clade which collectively form subgroup Dm. Finally genotypes E and F were not divided into subgroups because of their small numbers.

6.2.2 Database analysis

The complete nucleotide sequence of the 65 specimens was read into the fields of a database (BASICDB.DBF). HBV is circular and the number of nucleotides in the genome (3182-3221) is not divisible by three which results in the virus having an effective coding capacity of 9546 to 9663 nucleotides (3182 to 3221 amino acids). This is achieved by reading the code continuously in each of the three possible reading frames in turn (Galibert *et al,*. 1979). Thus programs were written to translate the viral sequence in all three of these reading frames and store the predicted proteins of the four viral ORFs in three protein database files (PROT1.DBF, PROT2.DBF, PROT3.DBF) labeled and numbered from the beginning of the reading frame in question (RF1, RF2, RF3). Figure 27 outlines the flow of the various in-house programmes together with the databases they use as input or generate as output. Appendix C gives the structure of the databases and the programme code. All programmes were written in dBase III+ (Ashton Tate). All graphs were drawn by importing the dBase III+ files into QUATTRO PRO (Corel 8, Perfect Office Suite).

6.2.2.1 Nucleotide and amino acid variation

The frequency of each nucleotide and amino acid at each position in the genome could be counted using the programmes SEQCT.PRG and AACT.PRG, respectively. The programmes GETMAXNT.PRG and GETMAXAA.PRG used the databases output by the former programmes (SEQCT.DBF, RF1CT.DBF, RF2CT.DBF and RF3CT.DBF) and identified the most common nucleotide, or amino acid, respectively, and its frequency at each position of the genome. The consensus nucleotide and amino acid sequences were then stored in appropriate databases SEQMAX.DBF, AA1MAX.DBF, AA2MAX.DBF and AA3MAX.DBF. Using these files, the frequency of each consensus nucleotide (or amino acid) at each position of the genome could be displayed graphically.

6.2.2.2 Rapid alignment

FILLAA.PRG compared each protein sequence to the consensus sequence and then aligned and listed the sequences. Aligned sequences were stored in the three database files AALIST1.DBF, AALIST2.DBF and AALIST3.DBF.

6.3 Results

6.3.1 The frequency and position of "hotspots"

The graph (Figure 28) shows the percentage frequency of each consensus nucleotide of the

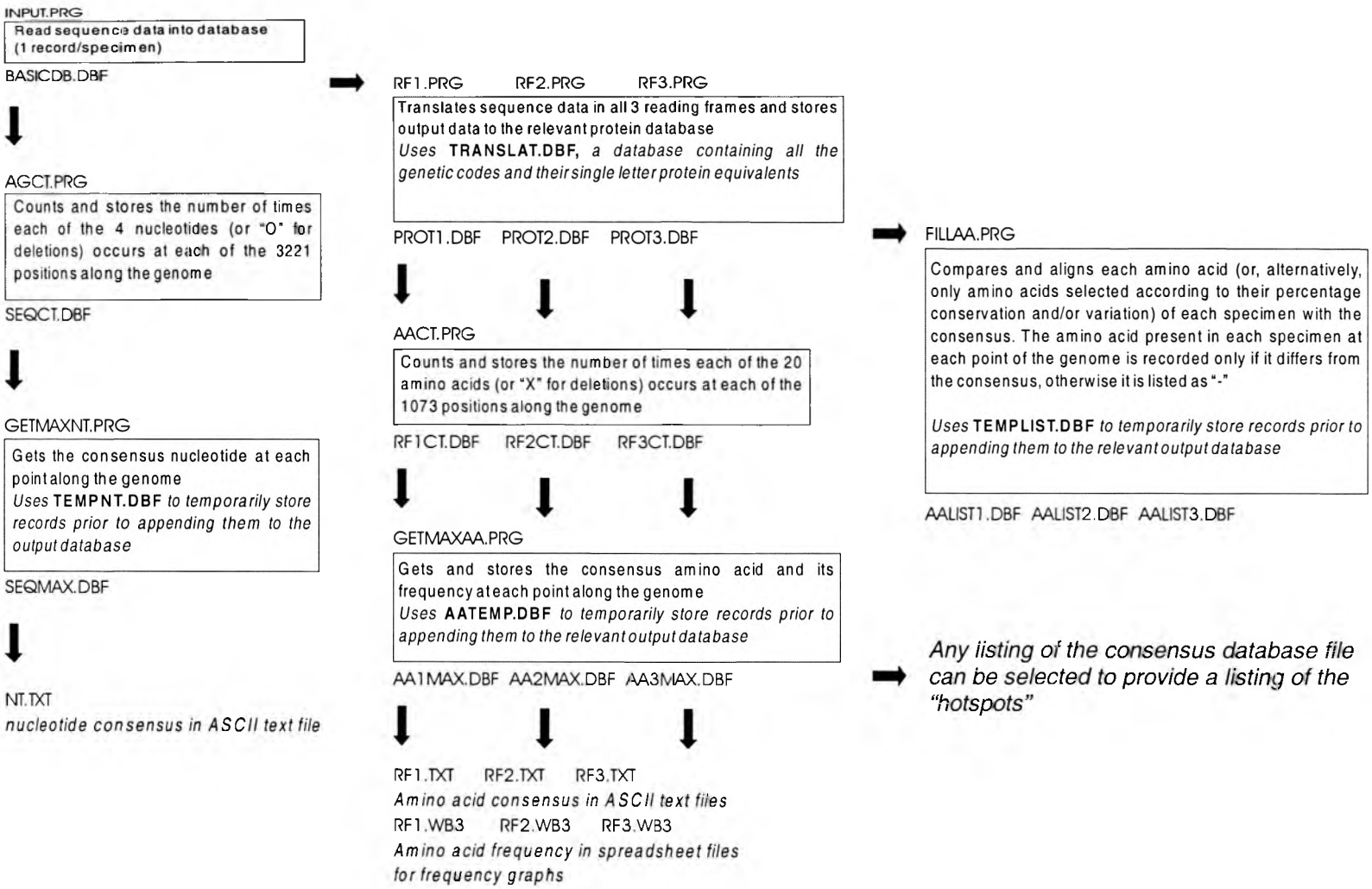


Figure 27. Flowchart of the computer programmes written to analyse the 65 full-genome sequences from GenBank together with the filenames of the databases they use as input or generate as output

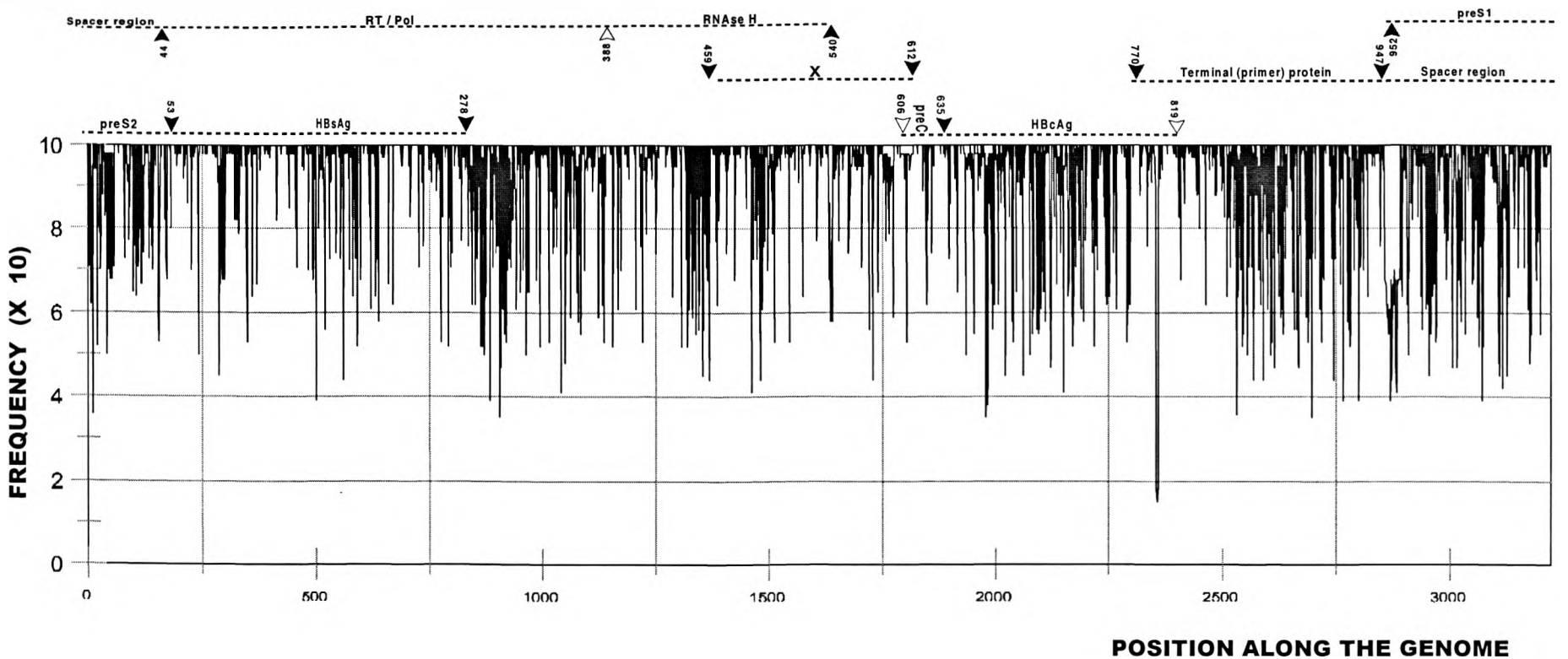


Figure 28. Frequency of the consensus nucleotide of 65 full-genome HBV sequences from GenBank plotted at each position in the genome
The lines above the graph show the position of the viral ORFs within each of the three nucleotide reading frames. RF1 is closest to the graph, RF2 is in the middle and RF3 is at the top. Numbers show the beginning and end of the various protein domains and all numbering is from the start of the RF in question

65 sequences at each base position along the genome. The depth of the troughs in the graph is proportional to the amount of variation at that particular site. The most conserved regions at the nucleotide level are found in the region where the HBsAg gene overlaps the RT/*Pol* domain of P, the latter part of the X gene where it overlaps the precore region of C and the beginning of the TP domain of P where it overlaps the end of HBcAg. The core gene shows considerable change at the nucleotide level. The preS1 and preS2 genes are also extremely variable at the nucleotide level.

Figures 29(a) to (c) show the percentage frequency of each of the amino acids, in each reading frame, plotted against their position in the genome. Coding regions within each translation frame are shaded in these graphs and can be seen to have much less variation than non-coding regions within the translation frame. This same pattern of uninterrupted troughs is seen in the non-coding spacer region (at the end of the second reading frame (RF2) and the beginning of RF3) between the TP and the RT/*Pol* coding regions of the P gene. The most extensive region of overlap between coding regions in different reading frames occurs between the surface antigen and the RT/*Pol* domain. The most conserved stretches of the genome at the amino acid level occur in this overlapped region of RF3.

6.3.1.1 "Hotspots" within the coding regions of RF1

There is considerable variation in the preS2 domain with only a short stretch of 8 conserved amino acids from amino acid 20-27 (Figure 30). Non-embedded parts of HBsAg are marked by stretches of deep troughs. These stretches are seen at the amino-terminal of the HBsAg domain (the first 8 amino acids, amino acid 52-60, numbering from the start of RF1) and within the intracellular (IC, amino acid 81-131) and extracellular (EC, 151-212) loops. The main variation is in the EC loop. The helices are relatively better conserved with variant amino acids tending to be at the start or end of the helix with the exception of the fourth helix which is fairly variable throughout. While known linear epitopes surround the EC region (HBs-150-158 precedes EC while HBs-210-224 follows it), none of the known linear epitopes fall within the loop itself.

Genotype- and subgroup-definitive change in the preS2 domain (Table 26, Appendix D; columns A-F RF1 amino acid 2-52) ranges between 11 definitive sites for subgroup D, which is most similar to the consensus, and 40 for genotype F. Calculating the number of definitive

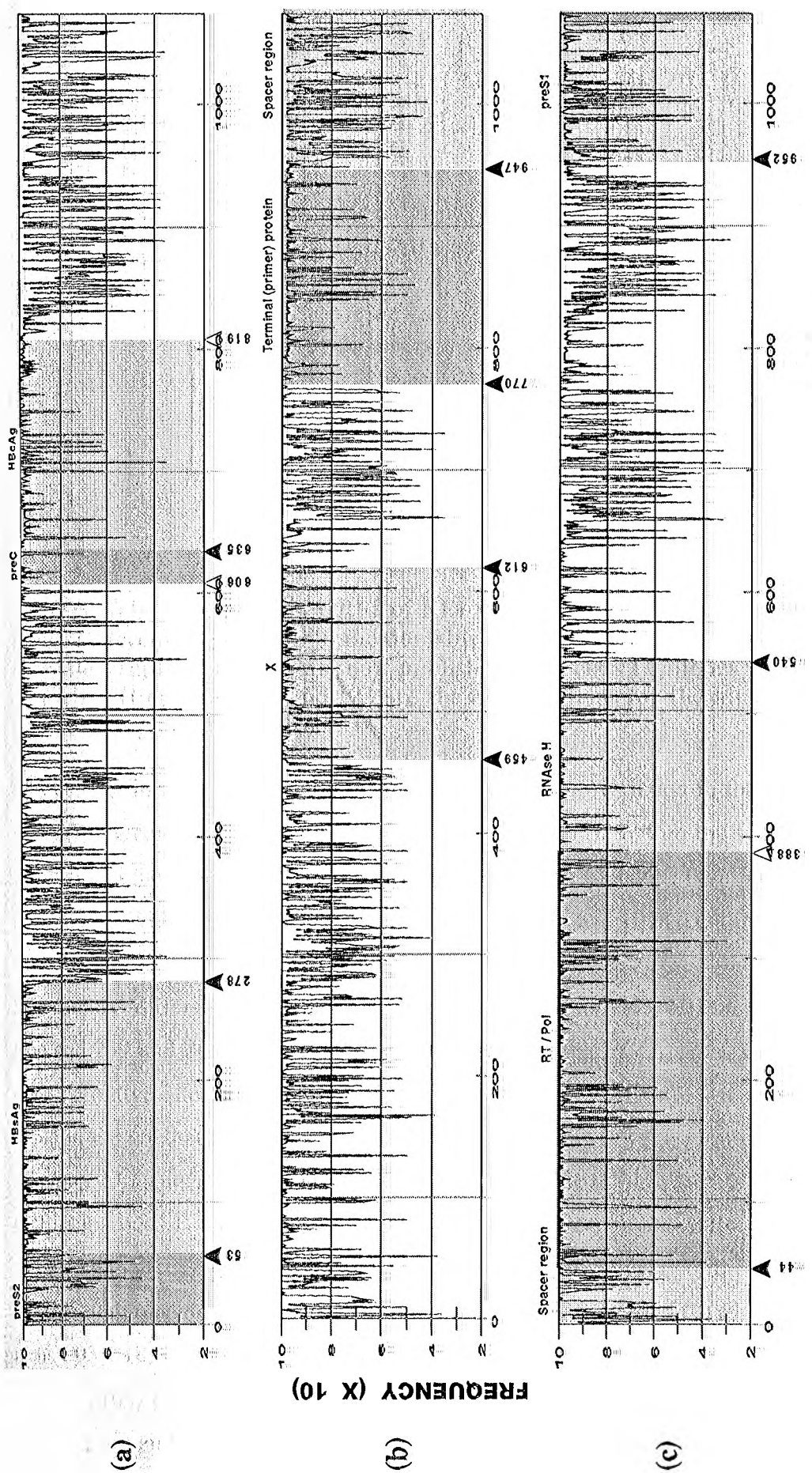


Figure 29. Frequency of the consensus amino acid of 65 full-genome HBV sequences from GenBank plotted at each position in the HBV genome in (a) RF1 (b) RF2 and (c) RF3. Shaded areas show coding regions within the 3 reading frames. Arrows show the start and end of the HBV protein domains

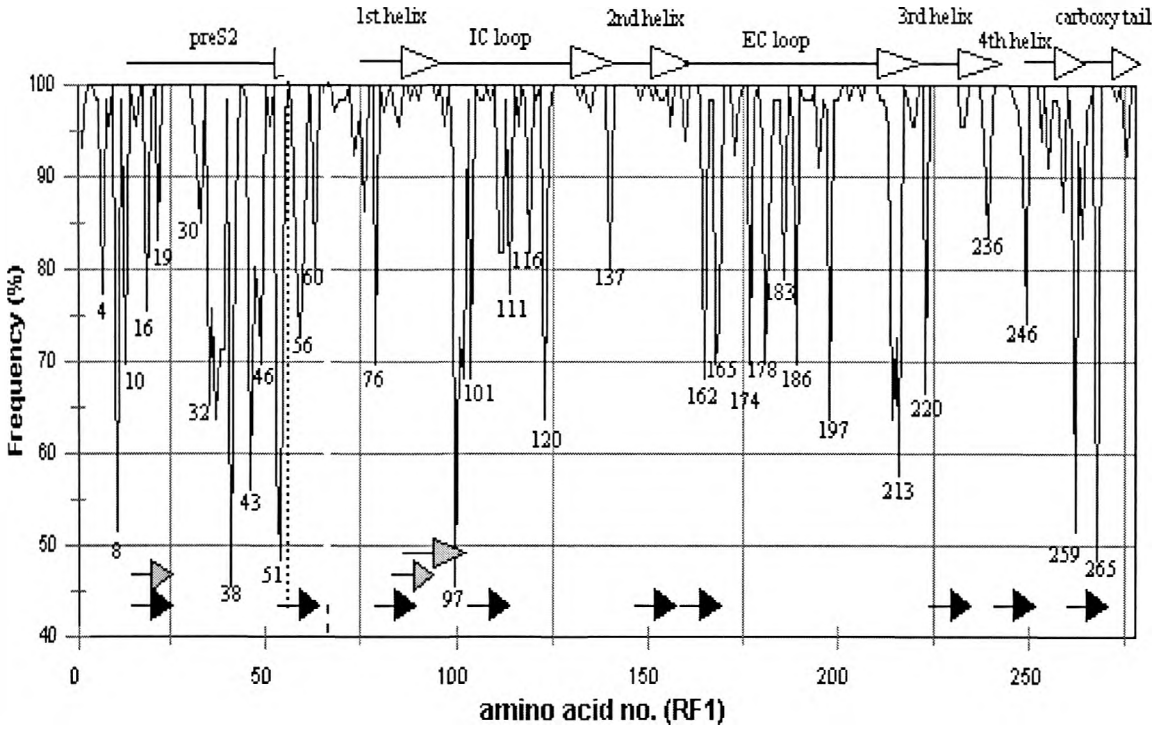


Figure 30. The "hotspots" in ORF1 of RF1 which codes for the preS2 and HBsAg domains. Different regions within the domains are marked by white arrows and labelled above the graph. Black arrows show published HLA class I epitopes while grey arrows show class II epitopes. IC and EC are the abbreviations for the intracellular and extracellular regions, respectively

sites per site in each functional domain confirms that the variation in the preS2 region (40 percent) is greater than in the HBsAg region (22 percent). Within the HBsAg, the amino-terminal and the fourth helix have the most change with 62 percent and 40 percent change, respectively, with helix two and three having the least (5 and 10 percent, respectively).

Subgroup A' differs from subgroup A at the sites 19, 29, 32, 44, 50 and 51 in the preS2 region and 259 and 261 in the HBsAg coding region (Table 26, Appendix D). These are the same definitive sites recorded in Chapter 5 except that two sites, 7 and 45, which were found in the southern African subgroup A' specimens are not present in the two published subgroup A' sequences. Genotype B shows considerable definitive change over this entire ORF. Subgroups B and B' are identical over the preS2 region but differ at amino acids 73 and 252 in the HBsAg region. In contrast, B and B" are identical over the HBsAg region but differ at amino acids 30, 32 and 38 in the preS2 region. Subgroup C (which contains specimens C03-C18) is a single subgroup with 12 definitive sites (amino acids 10, 55, 56, 97, 99, 101, 162, 165, 178, 212, 236 and 265), only one of these is within the preS2 region. Amino acids 35, 55, and 236 define minor subgroup partitions within subgroup C. The two sub-

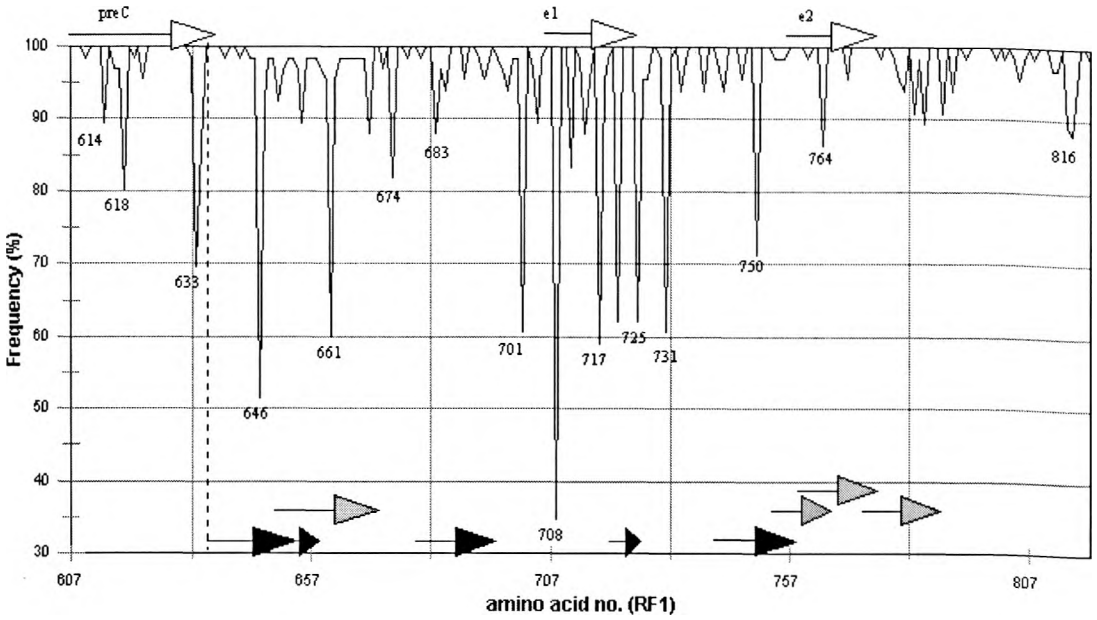


Figure 31. The "hotspots" in ORF2 of RF1 coding for the precore and core domains. Again different regions are shown by white arrows. e1 is exposed on the capsid surface and is an antigenic determinant of the assembled core while e2 involves portions of the protein which are buried and are only reactive in denatured core. Black and grey arrows show HLA class I and II epitopes, respectively

group Cq⁻ specimens (specimens Cq⁻01 and Cq⁻02) differ from subgroup C at positions 6, 38, 50, and 51 in the preS2 region and positions 55, 99, 211, 229 and 236 in the HBsAg domain. In this ORF specimens from genotype D share 11 common definitive sites (8, 46, 51, 98, 120, 174, 186, 211, 220, 259 and 265) but the Dm outliers have a single additional site, amino acid 39, which distinguishes them from subgroup D. The two genotype E specimens E01 and E02 vary from each other at only three amino acids (111, 177 and 259) but partition as a separate group on the basis of 12 definitive sites: 38, 46, 55, 76, 97, 174, 179, 192, 211, 220, 236 and 265. The 3 genotype F specimens have a common amino acid (different from the consensus) at 36 sites (10, 16, 29*, 32, 35, 38, 39*, 45, 48*, 51*, 60, 70*, 71*, 76, 97*, 99, 101, 113*, 120, 166, 179, 192, 210*, 211, 213, 220, 230*, 235*, 245*, 250*, 256*, 258*, 259, 265, 272*, 273*, 277*) and 18 of these sites (those marked with an asterisk) are unique to genotype F. Sites 54, 65 and 178 separate F02 from the other members of the genotype while F1 is only variant at positions 4 and 43, and F3 at 162 and 254.

The main variation of the precore and core region in RF1 is in the e1 region (Figure 31; amino acid 709-722). The embedded e2 region is well conserved. There is much less variation in this *gag* coding region than in the first, *env*, coding region of RF1.

The second ORF of RF1 is considerably shorter than the first (212 amino acids versus 277) and has far less definitive change, at the amino acid level (Table 26, Appendix D; RF1 amino acid 610-817). A two amino acid insertion (amino acids 787 and 788) is present in the 11 genotype A specimens only. Genotype A is well conserved in this reading frame with the exception of specimen A'02 which has a single nucleotide frame shift deletion at nucleotide 1931 which affects codon 644 and the following 21 codons introducing a non-sense mutation at codon 665 which would result in a truncated core gene. Specimen A09 has a guanine (G) to adenine (A) mutation at nucleotide 1897 which results in a tryptophan (TGG) to nonsense (TGA) change at amino acid 28 of the precore which affects translation of the precore transcript, although not the core transcript. The more common G → A mutation at nucleotide 1896 which results in a nonsense (TAG) codon at codon 28 is the consensus for both subgroups B and D. A total of 18 specimens in the database have this mutation.

Except for the two codon deletion at codon 787, there are no definitive sites common to all three subgroups of genotype B. Subgroup B has 9 definitive sites (614, 633, 646, 669, 708, 711, 717, 721 and 817) including the precore stop which is present in four (B05-06, B08 and B09) of the 6 members of this subgroup. Apart from the two deletions, there are only three definitive sites common to the subgroups B' and B" and these are identical to the definitive sites of subtype C. Subtype Cq-01 has the same three definitive sites (amino acids 661, 701 and 725) in common with subgroup C, together with one unique definitive site at amino acid 693. Subtype D differs from Dm in that it also has a 50 percent prevalence of the precore stop at amino acid 633 (D04, D06-D10) and an asparagine (Asn) at amino acid 674 (D02-D07). Another five definitive sites (not including the two amino acid deletion) are common to genotype D (646, 708, 717, 731 and 750). Most change is observed in genotype E (646, 674, 708, 717, 731, 750, 780, 787, 788, 815, 816) and F (650, 674, 683, 685, 704, 708, 715, 717, 721, 787, 788, 815, 816). Overall the region has 28 definitive sites in the total of 212 amino acids, giving it a 13.2 percent variability between the genotypes.

6.3.1.2 "Hotspots" within the coding regions of RF2

The X gene is fairly well conserved and change is concentrated between amino acids 488

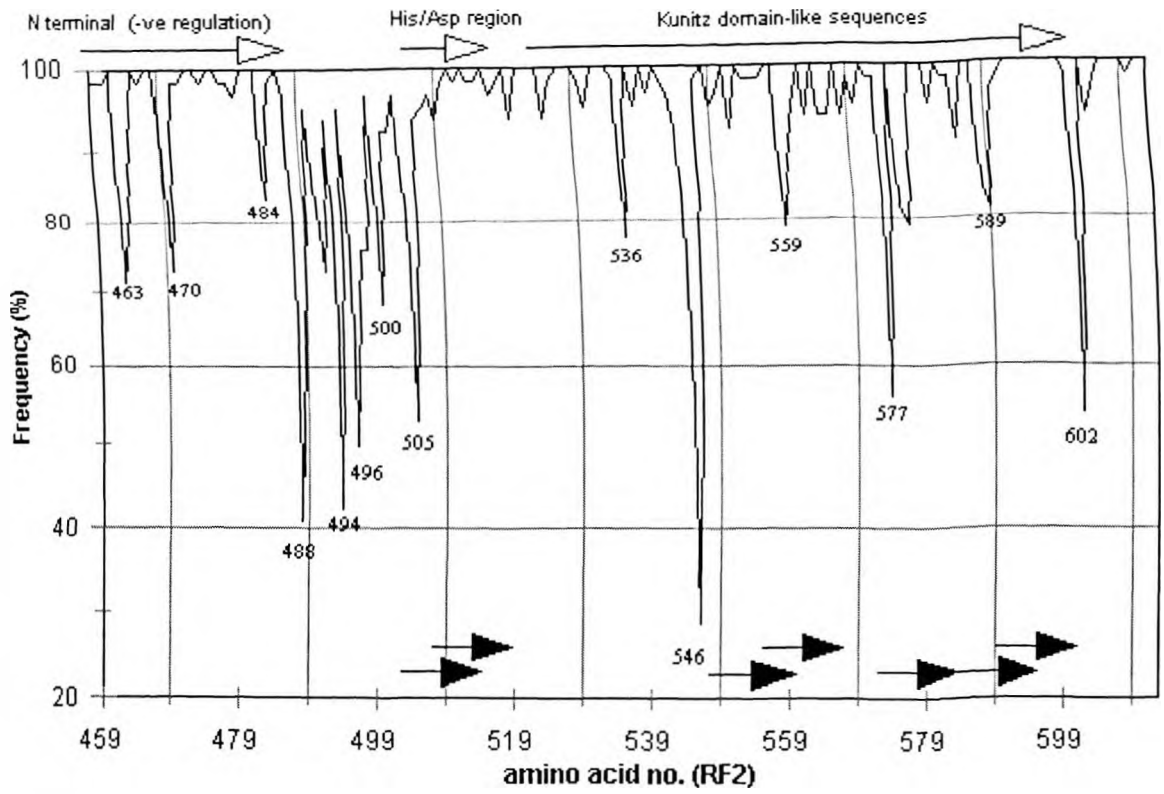


Figure 32. The "hotspots" on the X protein in RF2. The white arrows show the functional domains of X while the black arrows show the position of published HLA class I epitopes

and 505, in the region between the functional domains at the amino-terminal and the Kunitz domain-like sequences at the carboxy-terminal of the protein (Figure 32).

Both subgroups of genotype A have the same five definitive sites (464, 470, 496, 536 and 559) in the X coding region, ORF1 of RF2 (Table 26, Appendix D; RF2 amino acid 459-610). The subgroups of genotype B have 10 definitive sites (489, 494, 496-498, 505, 545, 546, 577 and 602). Subgroup B' shares a site at amino acid 501 with B" and site 576 with B and has a unique site at amino acid 502. Subgroup B" has a unique site at amino acid 576. Subgroup C has three unshared definitive sites (494, 574 and 602) and a further 6 definitive sites which it shares with subgroup Cq' (463, 488, 492, 496, 500 and 546). Cq' also has three sites which differentiate it from subgroup C (470, 506 and 545). Genotype D specimens share four sites (488, 496, 505 and 574). Subgroup Dm alone has a unique site 491 while subgroup D has three additional unique sites (494, 504 and 546). Seven sites define genotype E (480, 484, 488, 491, 494, 496, and 574) while 22 sites (463, 484, 487, 489, 494, 495, 497, 498, 509, 520, 544, 545, 546, 549, 560, 561, 563, 565, 566, 568, 570 and 602) define

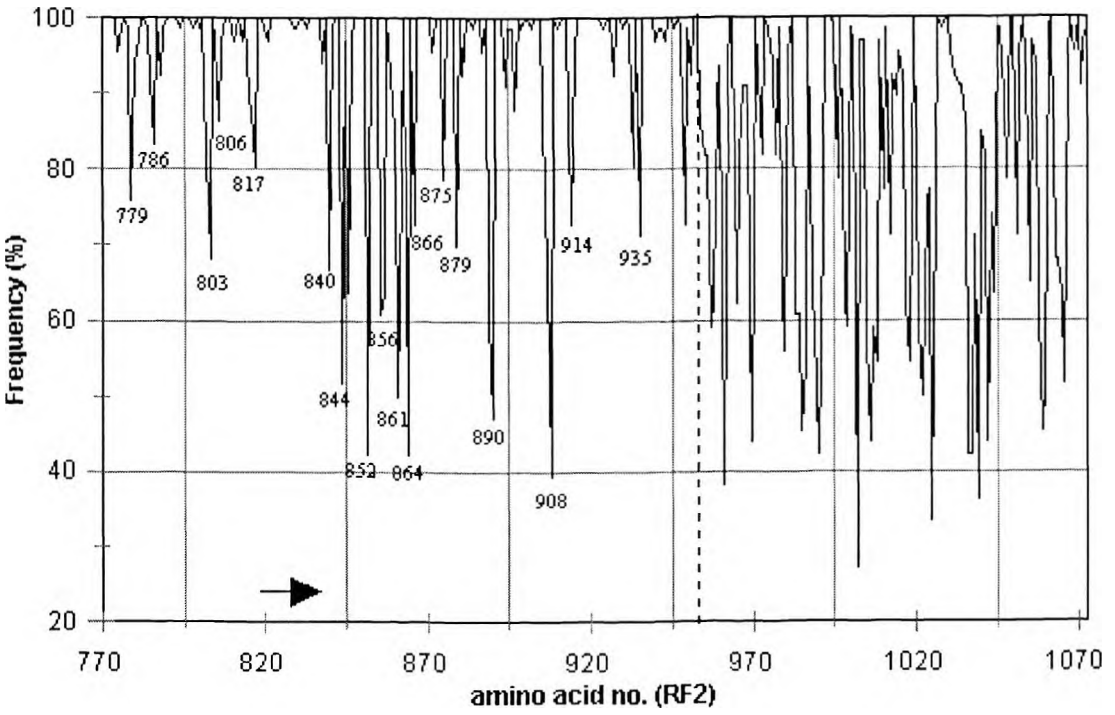


Figure 33. The "hotspots" on the TP domain of the P gene in RF2 including the spacer region in this reading frame. The black arrow shows the single published class I epitope in this region

genotype F. The percentage of definitive sites in this reading frame is 25.3 percent although, without the sites specific for genotype F only, this drops to 16.2 percent.

The TP domain is fairly variable throughout without long conserved stretches as can be seen in the P coding regions in RF3 (compare Figures 33 and 34). The first 65 amino acids of the TP, which are exclusively involved with priming, are more conserved than the rest of the protein which is involved in protein-protein binding as well (Table 26, Appendix D; RF2 amino acid 775-948). The start of the spacer region sequence at amino acid 948 (Figure 33) contrasts sharply with the coding region preceding it.

Genotype A is defined by the two amino acid insertion at sites 786 and 787 as well as 8 other common sites (802, 803, 840, 852, 889, 890, 906 and 907). Genotype B has only four common sites (852, 864, 875 and 949) with a further two sites common to subgroups B and B' only (860 and 861), and only one site common to subtype B and B" alone (amino acid 933). Subtypes B'(785, 840 and 907) and B (806, 817 and 935) have three unique sites each while subtype B" is only unique at amino acid 840. Genotype C is defined by four sites

(866, 890, 908 and 914) while Cq⁻ has one unique site at amino acid 817. Subtype D has 11 definitive sites (779, 844, 846, 856, 857, 861, 862, 864, 879, 908 and 935) in addition to the deletion at 786/787. Subgroup D has one unique site at amino acid 803 while subgroup Dm has two unique sites at 889 and 949. Genotypes E (779, 780, 786, 787, 802, 815, 816, 817, 838, 844, 846, 852, 856, 857, 862, 864, 879, 881, 887, 889, 890, 897, 906, 907) and F (775, 786, 787, 815, 816, 817, 840, 844, 846, 852, 856, 857, 859, 861, 862, 864, 872, 875, 881, 889, 894, 897, 908, 933) are extremely different from the consensus in this ORF each having 24 definitive changes. Total definitive change in this RF is 22.8 percent.

6.3.1.3 "Hotspots" within the coding regions of RF3

The P gene is continuous reading from RF2 into RF3, although the first 43 amino acids in RF3 are still part of the spacer region. The RT domain consists of three well conserved areas sandwiched between three more variable regions (Figure 34). Most change is between amino acid 134-196 and 264-314. The active site of the RT domain, amino acid 246-249, which contains the residues YMDD is 100 percent conserved in my database. The genotypes are well defined in the RT/*Pol* domain (Table 26, Appendix D; RF3 amino acid 44-386). Genotype A is defined by 13 sites (50, 96, 146, 165, 167, 169, 182, 194, 196, 306, 314, 375 and 379) and a further three sites (206, 260 and 296) differentiate subgroup A from subgroup A' while one site (amino acid 172) defines subgroup A' alone. Only five sites are common to all three genotype B subsets (56, 59, 81, 194 and 265). A further three sites are common to subgroups B and B' only (177, 281 and 375) while one site (170) is common to subgroups B' and B". Four sites are unique to subgroup B (174, 196, 314, 380) and B' alone (50, 167, 177 and 314) and three sites are unique to subgroup B" (119, 281 and 310). Eleven sites define genotype C (56, 81, 96, 152, 161, 164, 167, 174, 266, 267 and 310) while five sites are only present in subgroup C (98, 188, 264, 312 and 360). Subgroup Cq⁻ does not agree with the consensus or subgroup C at amino acid 134. Ten sites define genotype D (134, 188, 192, 264, 299, 300, 302, 309, 314 and 360) but the subgroup D differs from subgroup Dm at sites 97, 165, 291 and 306 and subgroup Dm differs from D at an additional two site (167 and 173). Genotype E is defined by 12 sites (50, 54, 96, 97, 134, 172, 173, 191, 207, 302, 309 and 360) and genotype F by 27 sites (50, 54, 59, 81, 96, 134, 161, 165, 166, 177, 188, 191, 192, 194, 250, 274, 280, 281, 289, 291, 306, 310, 314, 319, 359, 365 and 379). Total variation over this essential stretch of P, the RT domain, is 16.2 percent.

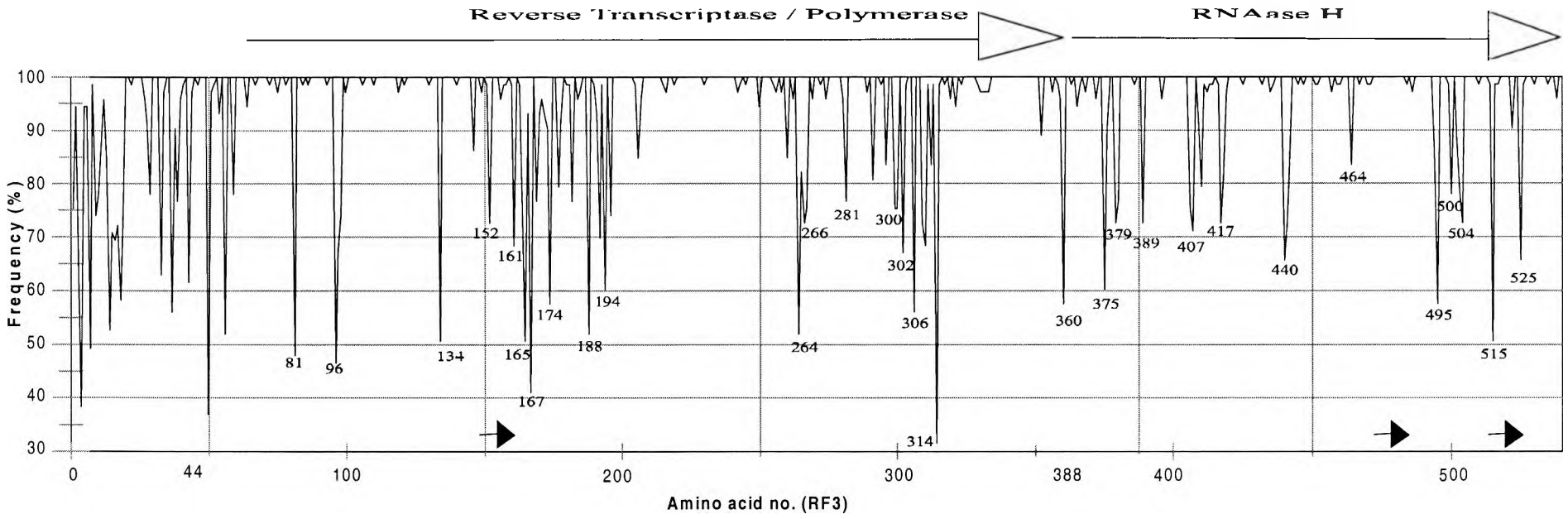


Figure 34. The "hotspots" on the RT/*Pol* and RNAse H domains (demarcated by white arrows) of the P gene in RF3. The black arrows show the published HLA class I epitopes in this region

The RNAse H domain is well conserved particularly in its middle region from about amino acid 418-494. The RNAse H region is relatively short, 152 amino acids, and reasonably well conserved (Table 26, Appendix D; RF3 amino acid 389-538). Only two sites are common to genotype A (464 and 525) with a further two which are unique to subtype A (495 and 504) and one which is unique to subtype A' (503). Genotype B has three sites common to all members of the genotype (418, 503 and 515). One site is common to the subgroups B and B" (522) while subtype B' has a unique site at amino acid 502. Genotype C is defined by a single site (389) while subtype C alone is defined by additional sites at 410 and 500, while site 502 is unique to subtype Cq . Four sites define genotype D (407, 440, 441 and 515) and another four sites are unique to subtype D alone (406, 417, 442 and 494). At two of these sites (417 and 495) subtype Dm has a codon different to both subtype D and the consensus sequence. Three (409, 494 and 504) and four (440, 486, 504 and 515) sites define genotype E and F, respectively. The overall variation over this domain is only 13.8 percent (Table 26, Appendix D; RF3 amino acid 389-538).

At the start of the spacer region of the P gene there is a definitive 11 amino acid deletion in genotype D specimens only, and a single amino acid deletion in genotype E specimens.

Like the preS2 region, the preS1 region shows considerable heterogeneity with only short stretches of conserved sequence (Figure 35). The longest stretch of conserved sequence is

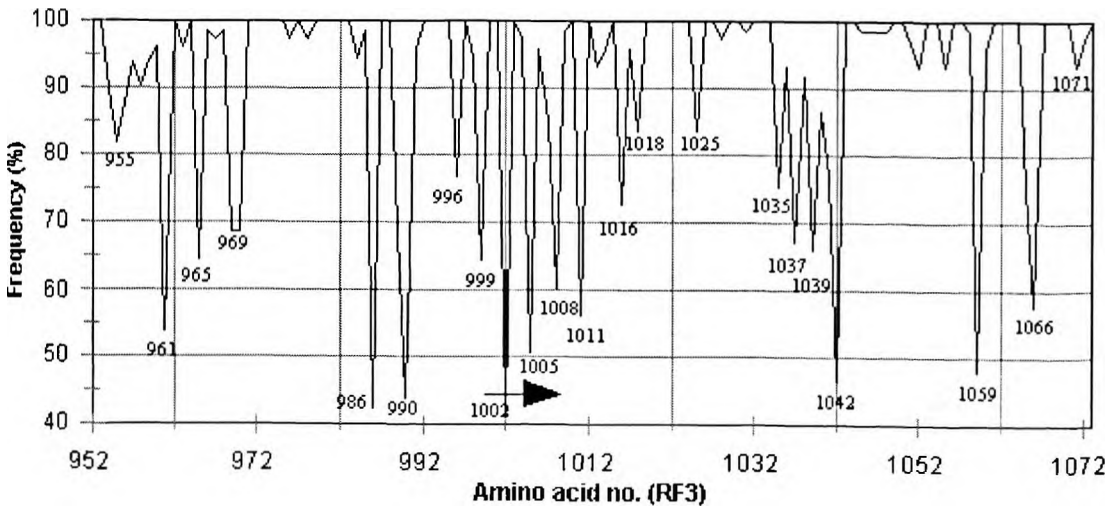


Figure 35. The "hotspots" on the preS1 domain of the S gene in RF3. The black arrow shows the single published HLA class I epitope in this region

from amino acid 971-983. Although quite heterogenous between specimens, the preS1 region, more than any other region, shows predominantly definitive change being well conserved between isolates belonging to the same subtype (Table 26, Appendix D; RF3 amino acid 953-1066). Eight sites define genotype A (986, 999, 1002, 1005, 1008, 1011, 1042 and 1066). Four sites define subtype A only (1018, 1025, 1040 and 1041) while two sites define subtype A' only (1025 and 1037). Five sites are common to the subtypes of genotype B (986, 996, 999, 1011 and 1059). Subtype B" has 6 unique sites (961, 990, 1006, 1041, 1052 and 1055) while subtype B' has one (1035) and subtype B has two (1007 and 1038). Five sites define genotype C (986, 1002, 1005, 1008 and 1066) while one site defines subtype C alone (961). There are no sites unique to only one of the subgroups of D in this region although 12 sites, in addition to the deletion of 11 amino acids which is unique to genotype D, define the genotype. Twenty-two sites, define the genotypes E (954, 955, 956, 957, 958, 961, 963, 965, 967, 969, 970, 989, 990, 996, 1002, 1005, 1011, 1035, 1037, 1039, 1042 and 1060) and F (954, 955, 956, 958, 961, 965, 978, 990, 991, 998, 1002, 1004, 1005, 1011, 1013, 1014, 1035, 1039, 1042 1051, 1055 and 1059), respectively.

6.3.2 Individual change within specific isolates

Individual, or non-definitive, change within specific isolates is listed in the right hand columns of Table 26 (Appendix D) and labelled value 2, value 3 etc. In addition, for each specimen and each HBV protein domain, I counted the number of times the specimen varied from the overall HBV consensus sequence. If the variation defined the genotype to which the specimen belonged it was added to the total of definitive changes otherwise it was added to the total of individual changes. Finally the two totals were plotted as a stacked bar (Figures 36-38) with the height of the grey portion of the bar representing the number of genotype specific, or definitive, variations within the specimen and the height of the black portion showing change in regions which do not define the genotype. Recombinants, outliers, specimens from different genotypes or wrongly typed specimens will then show a difference in the grey portion of their bar while the black portion shows the specimens undergoing the greatest amount of individual change. Domains in which the majority of specimens show considerable individual change are probably under the greatest selection pressure.

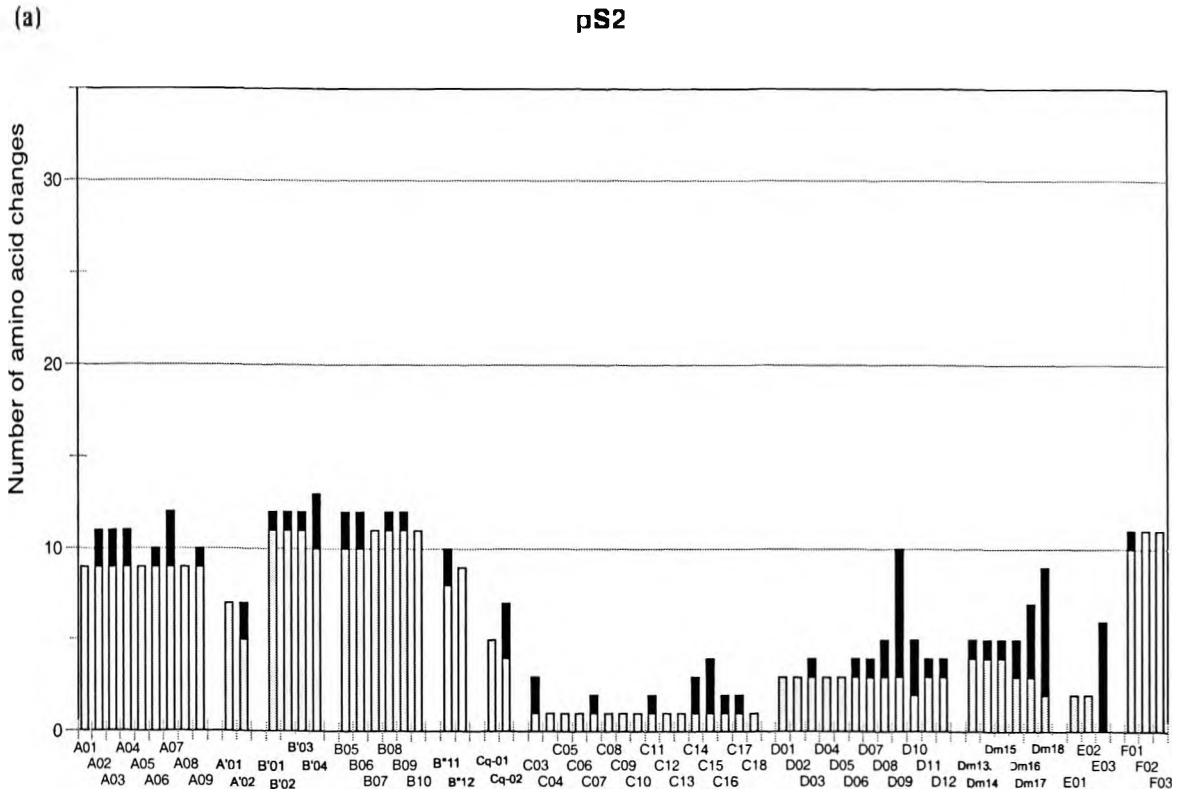


Figure 36. For legend see Page 108

6.3.2.1 Non-definitive change in RF1

The preS2 region of ORF1 in RF1 has predominantly definitive change (Figure 36a). D09 and Dm18 have the most non-definitive change with 7 changes each. Five of the Dm18 changes are consecutive genotype A definitive changes (amino acid 43, 44, 46, 50 and 51) and the last two changes are shared by Dm16. The variation from the genotype D consensus is also reflected in the grey portion of the bar which is shorter than that found in more typical members of the genotype. E03 has been grouped with genotype E in this series but is now known to be an isolate of chimpanzee hepatitis B virus. It varies from the consensus in 6 places in the preS2 region and shares no definitive sites with the other members of genotype E.

In the HBsAg region at the protein level variation is random non-definitive change with the exception of Dm16 (amino acid 166, 174, 183, 186, 211, 213, 220 and 246) and Dm18 (amino acid 246, 261 and 265) which have consecutive amino acid change which is mosaic (D/A) in nature. Apart from Dm16 and Dm18 (with a total of 14 and 9 non-definitive changes respectively), E03 (11 changes), B'04 (10 changes), C16 (7 changes) and C17 (9

(b) HBsAg

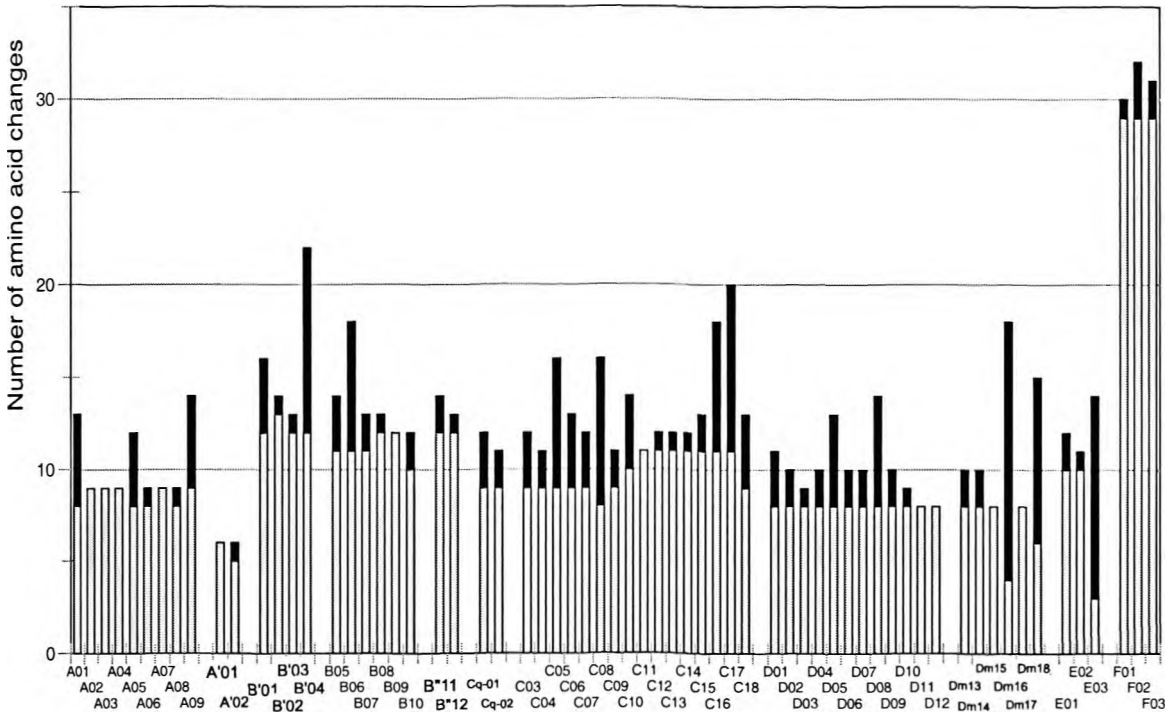


Figure 36. For legend see Page 108

changes) have the most individual change. B06 and C05 have 7 changes each while C08 has 8 changes, all other specimens have 6 (2.6 percent) or less non-definitive changes (Figure 36b).

Specimen change is less random in the second ORF of RF1, the precore/core region. While many specimens are well conserved, others show considerable non-definitive variation (Figure 36c). The most heterogenous isolates include B05 and B06 of genotype B with 17 changes each. Specimens with the most change from genotype D are D10 (15 changes), D04 (14 changes), D09 (12 changes), D07 and D06 (11 changes). The premature stop at codon 28 of the precore transcript is a common feature of these isolates and 9 other variants [D08 (7 changes), B'04 (8 changes), B08 (6 changes), C04 (five changes), C05 (five changes), C16 (11 changes), C17 (13 changes), C18 (9 changes), Dm17 (9 changes) and E03 (13 changes)]. Only three patients without the precore nonsense codon also have this level of variation and two of these have other nonsense mutations in the precore region. A'02, with 18 changes, has a mutation in the core gene which causes a frame shift within the core (amino acid 644-665) and then, after 21 variant amino acids, codes for a premature stop at

amino acid 665. Another genotype A specimen, A09 with 11 changes, is the only specimen in the database to have the nonsense codon TGA (as opposed to TAG) at site 28 of the precore. B'01 with 10 changes does not have a premature precore stop in this reading frame but does have the G → A mutation at nucleotide 1899 which changes the last codon in the precore region from a glycine (Gly) to an Asn. This latter mutation more often occurs in conjunction with the mutation at nucleotide 1896 and in my database all other specimens with the precore stop mutation have the Gly → Asn codon 29 switch as well. E03 differs in 6 positions from genotype E while E01 and E02 are identical over this ORF (Figure 36c).

6.3.2.2 Non-definitive change in RF2

ORF1 of RF2 codes the X gene and E03 has 23 individual changes from the consensus in this region. Both Dm17 and Dm18, have D/A mosaics in the X reading frame sharing amino acids 464, 470, 504, 505, 559, and 574 with genotype A. Dm18 has an additional genotype A site at amino acid 487. The two members of subtype A' have 11 and 9 changes, respec-

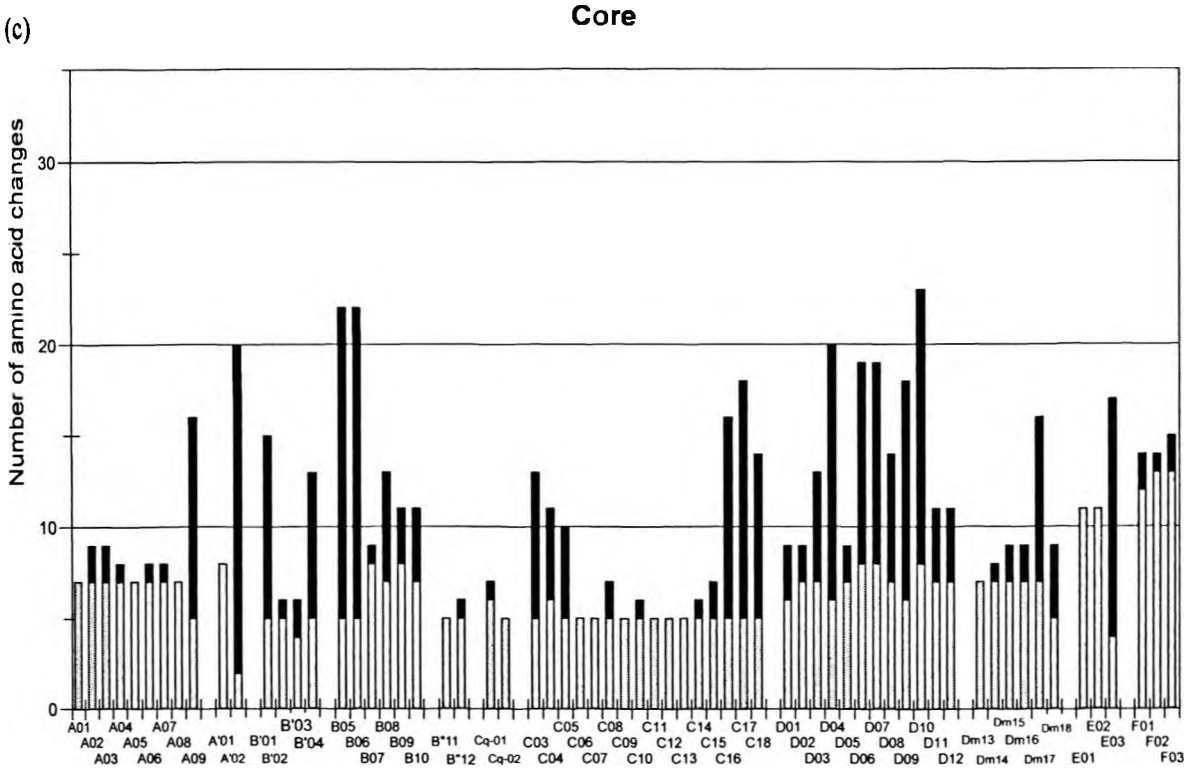


Figure 36. The number of definitive (grey) versus non-definitive (black) amino acid changes in RF1 coded regions of 65 specimens grouped by genotype in (a) the preS2 domain (b) the HBsAg region and (c) the precore/core region

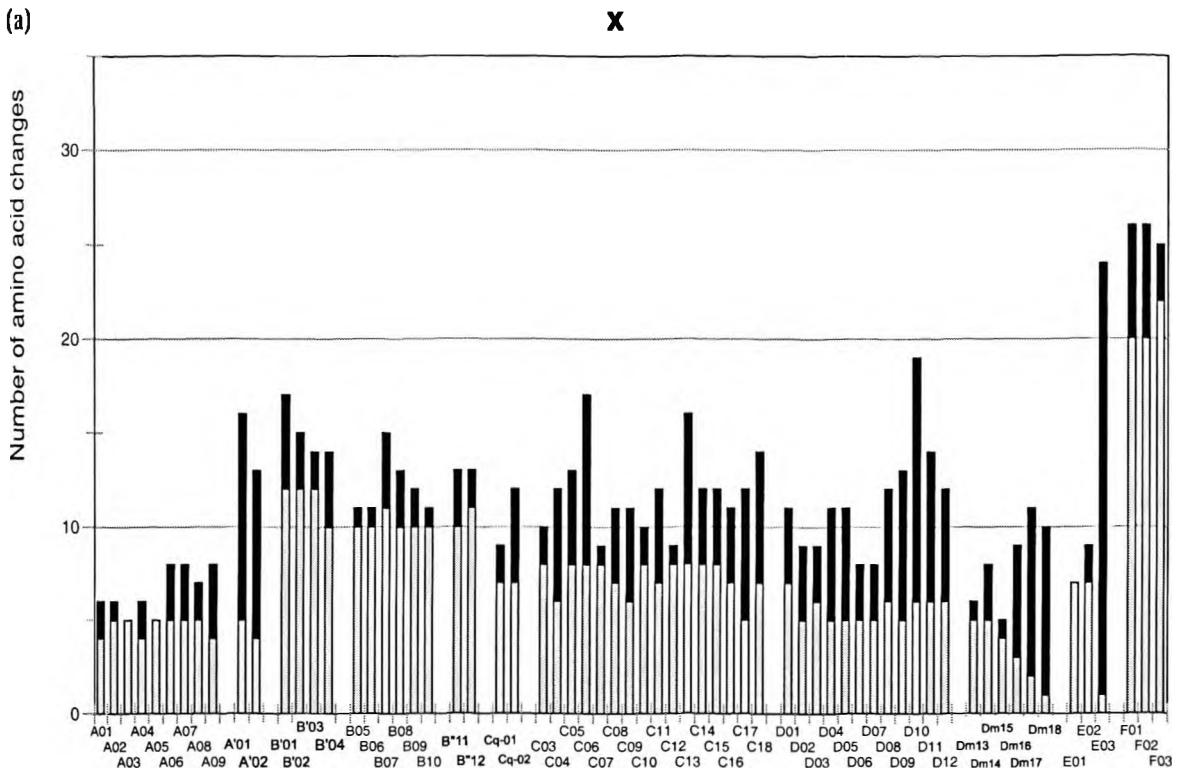


Figure 37. For legend see Page 110

tively, D10 has 13 changes and C06 has 9 changes. Specimens with changes to amino acids 588 and 599 in the core promoter region include: A06 (3), A07 (3), A09 (4), B'01 (5), C04 (6), C05 (5), C18 (7), D08 (6) and D10. C03 (2) has only the first mutation and C16 (4) and C17 (7) have only the second (Figure 37a).

Point mutation in the TP domain has been reported to be responsible for replication defects and in particular defects in the packaging of the pregenomic RNA (Blum *et al.,* 1991). Only specimen A09 has the Thr to Pro mutation reported to cause this defect. A'01 has 15 changes in this reading frame and E03 has 10. All other specimens have 8 (D10) changes or less (Figure 37b).

6.3.2.3 Non-definitive change in RF3

Long mosaic fragments of Dm16 (amino acid 165-260), Dm17 (274-540) and Dm18 (43-134 and 246-407) are present in the RT/*Pol* region of ORF1 in RF3. These three specimens have non-definitive change at 15, 12 and 21 sites, respectively. Apart from specimen E03 which has 20 changes in this reading frame, only B05 (10 changes), B06 (10 changes), C18

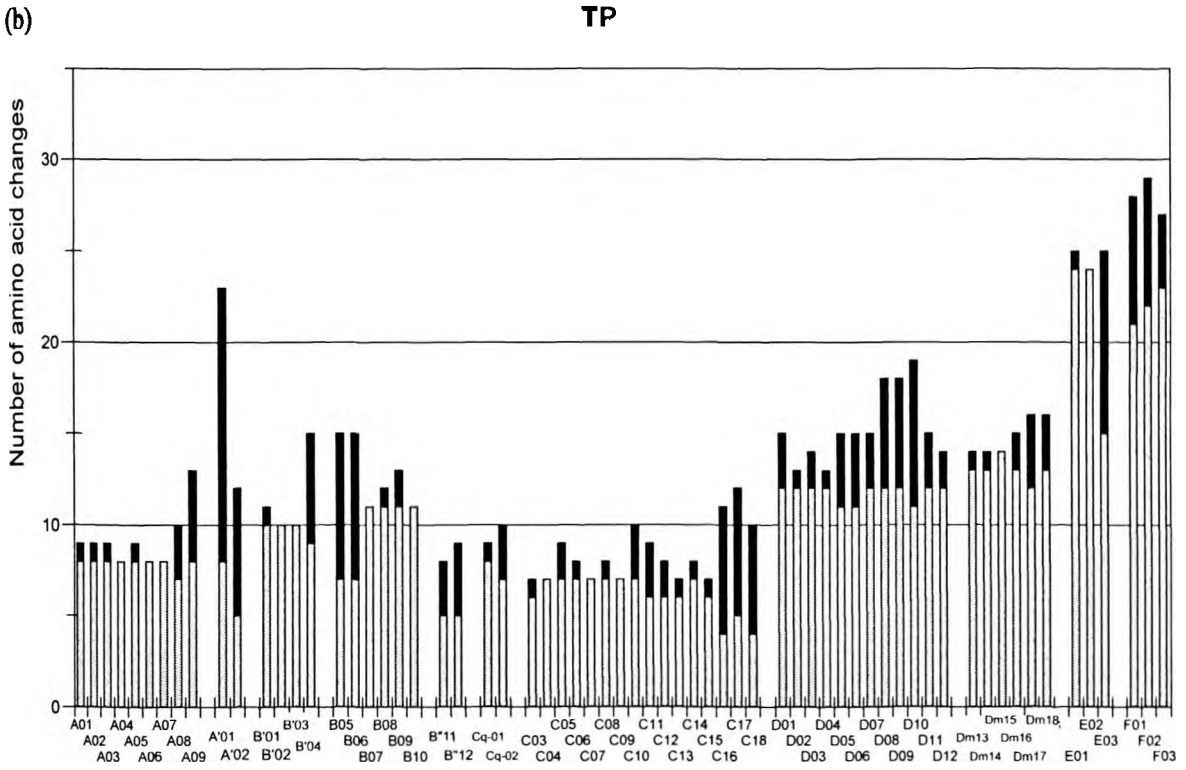


Figure 37. The number of definitive (grey) versus non-definitive (black) amino acid changes in RF2 coded regions of 65 specimens grouped by genotype in (a) the X region and (b) the TP domain of P

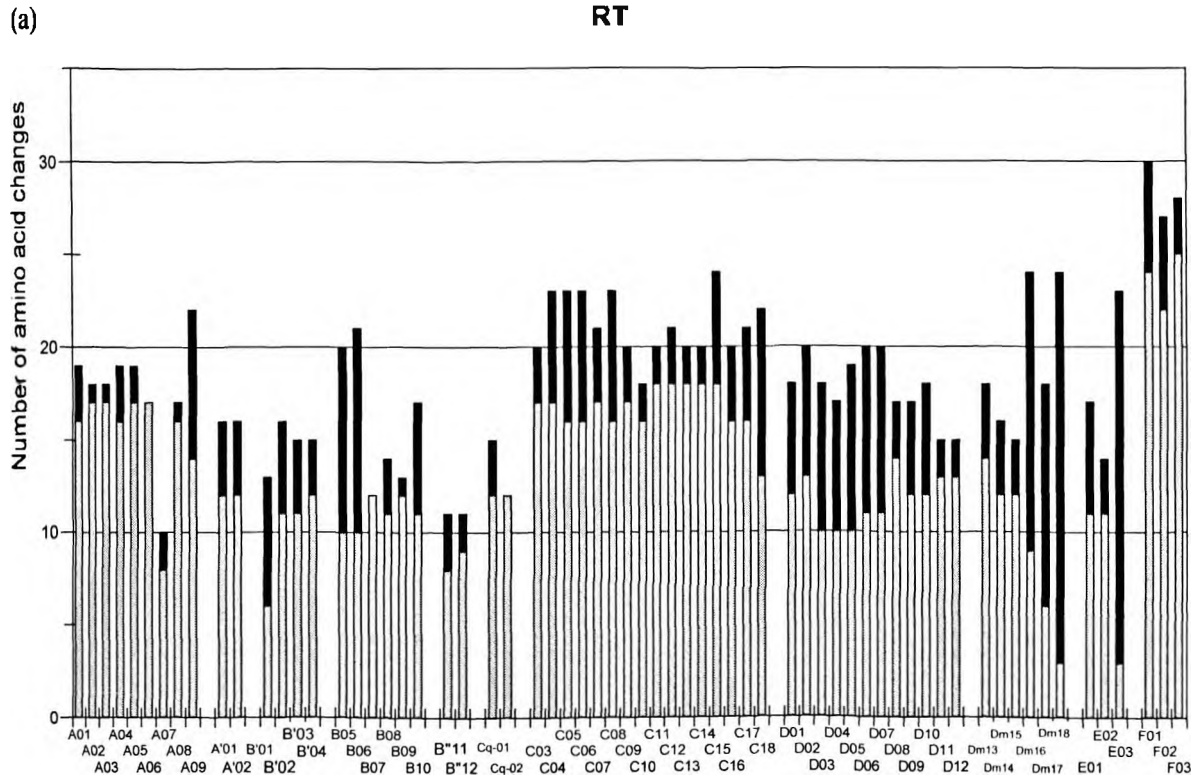


Figure 38. For legend see Page 112

(9 changes), D05-D07 (9 changes each), D03 and A09, with 8 changes each, have notable non-definitive change (Figure 38a).

The RNAse H domain which is continuous with the RT domain has even less non-definitive change. Dm17 has 11 changes and this is mosaic in nature and E03 has 8 changes. All other specimens have five changes (3.3 percent) or less (Figure 38b).

The preS1 region has the least non-definitive change of all ORFs (Figure 38c). With the exception of E03 which has 24 non-definitive changes, Dm16 with 8, B'01 with 6 and specimens D10, A01, A02 and C12 with five, all other specimens have three (2.5 percent) or less changes from the consensus of their subgroup.

6.4 Discussion

I plotted the frequency of the consensus nucleotide (or amino acid) at each point along the genome in a database of 65 full-genome specimens in order to locate the "hotspots" of variation in HBV. To distinguish between nucleotide or amino acid differences developed within

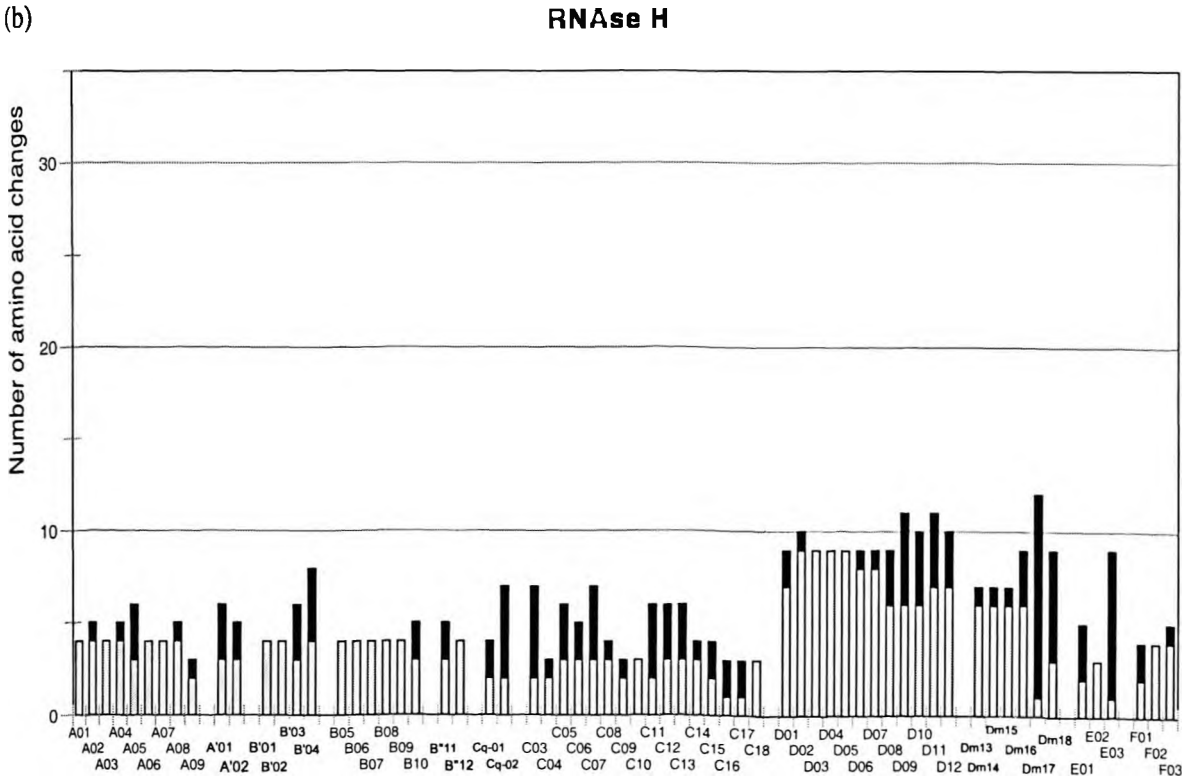


Figure 38. For legend see Page 112

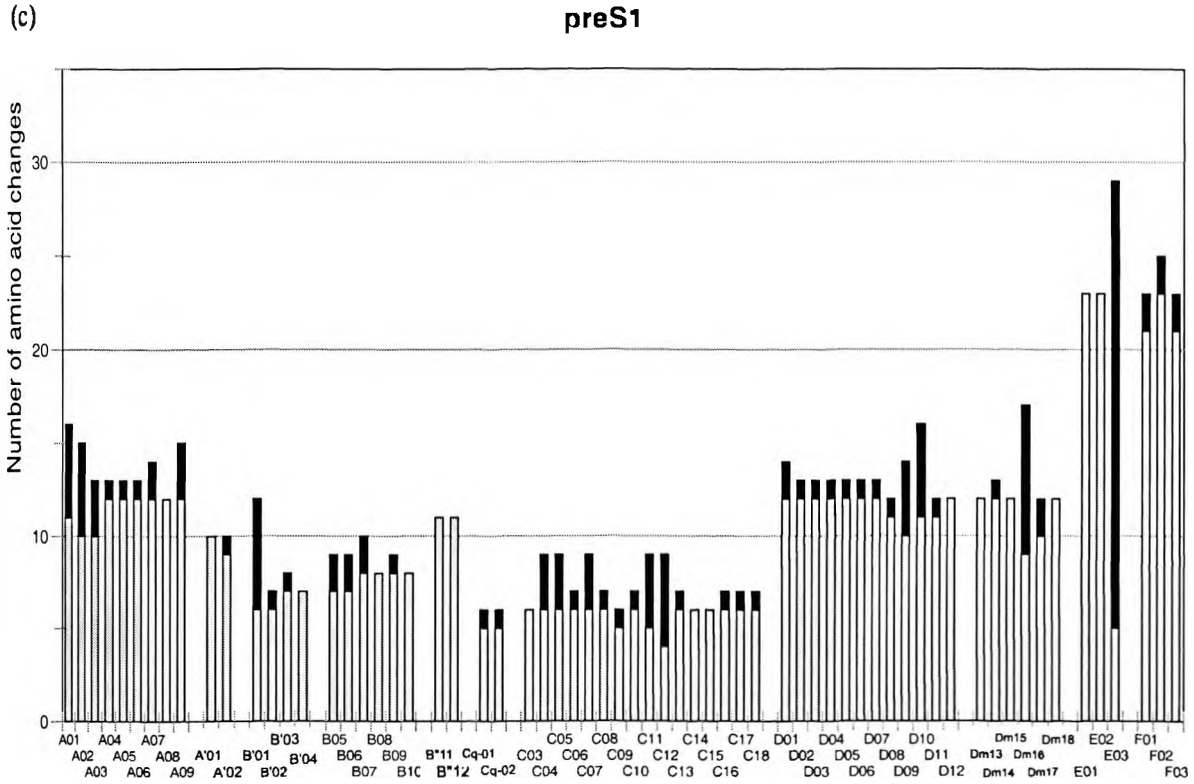


Figure 38. The number of definitive (grey) versus non-definitive (black) amino acid changes in RF3 coded regions of 65 specimens grouped by genotype in (a) the RT and (b) the RNAse H domain of *Pol* and (c) the preS1 domain of *env*

an individual and those inherent to the strain of the infecting virus I divided change into definitive and non-definitive change. The former referred to the consensus sequence which defined the 11 subgroups, while the latter referred to change within isolates which differed from the consensus of the subgroup to which it belonged. Since the 11 subgroups were defined on the basis of the full-genome phylogenetic tree of these 65 specimens, examining the definitive change within each partition, at the amino acid level, reveals the amino acid variation in each domain which is specific for the subtype. Using a simple database to store and manipulate the nucleotide data I was able to quickly translate and align any number of sequences and also to calculate the consensus of the complete database, or any subset thereof. The results of these manipulations provided baseline data which was also used in other studies.

While definitive change represents natural selection in the past, frequent change, not defined by or within the subgroups highlights current "hotspots" favoured by natural selection and can pinpoint important structural or control regions of the genome, T-cell contact sites

within B cell epitopes, or anchor residues within linear epitopes. One such site was found at amino acid 19 (RF1) where an Phe to Lys change occurred in 5/9 subgroup A specimens as well as in four genotype B specimens and one each from genotype C and D. This mutation had a 16.9 percent incidence in my database and was represented in five of the 11 subgroups. The site is within the preS2 B cell epitope and is one of four amino acids thought to be responsible for the universal "*b*" determinant of the preS2 region (Usuda *et al.*, 1999). Although the other three amino acids contributing to *b* are 100 percent conserved, the effect of the change in nearly 17 percent of specimens and 56 percent of genotype A specimens needs to be assessed, since the *b* epitope is used to determine whether specimens can in fact be typed by HBV genotype-specific EIA.

Mutations in the extracellular region of the HBsAg are responsible for serotype-specific change (Norder *et al.*, 1992). Amino acid 174 (RF1) is responsible for the *d/y* switch (Phe = *d* and Tyr = *y*). One specimen, C10, from genotype C, is known to be serotype *ayr*, and had a Tyr as expected at RF1 amino acid 174. Dm16 from the *ayw* exclusive genotype, genotype D, had the *d* determinant at this point, further evidence that the variation is the result of recombination. The chimpanzee specimen, E03, also had the *d* determinant. Amino acid 179 predicts the *w* epitope (Pro = *w1/w2*, not Pro = *w3* and Lys = *w4)*. There were 6 specimens from genotype D with the *w*3 phenotype indicated by an absence of Pro at amino acid 179. *w1* is distinguished from *w2* by the amino acid at site 186 (Tyr = *w1* and Phe = w2). With the exception of specimens Dm16, which has the *w2* specificity, all remaining members of genotype D had the *w1* determinant. All genotype A and B specimens in the database had serotype *adw2*. Okamoto *et al.* (1988) used *adw1*, *adw2* and *adw3* to describe the three subgroups of genotype B, but this study shows that his use of the sub-determinants *w*1-*w*3 is not based on the 3D epitopes of the HBsAg gene product.

Mosaic variation was found in specimens Dm16-Dm18 and in two of the three subgroups of genotype B in the core gene region. Dm16 has a non-genotype D serotype of *adw2* and Dm17 and Dm18 are distinct from the rest of genotype D in the preS2 region based on the epitopes of Usuda *et al.* (1999). This phenomenon is studied further in Chapter 7.

E03 was extracted from the serum of a naturally infected chimpanzee, as opposed to A07 and C18 which were extracted from serum of laboratory infected chimpanzees. It has been

speculated that this isolate represents adaptation of the virus to the chimpanzee host. This specimen lacks definitive sites of any of the 6 genotypes and has considerable non-definitive variation not present in the other specimens of the database. The discovery of several chimpanzee hepatitis virus isolates from chimpanzees infected in the wild (Hu *et al.*, 2000; Takahashi *et al.*, 2000) has recently confirmed that E03 is an isolate of the chimpanzee hepatitis virus and not a human HBV as previously thought.

A high rate of mutations in the core gene during the immune clearance phase of chronic hepatitis B virus infection is associated with the presence of the precore stop mutation at nucleotide 1896 which affects codon 28. The appearance of the precore stop codon mutation has been significantly associated with HBeAg negativity and active liver disease, and is thought to affect the immune clearance of HBV since the HBeAg is thought to induce viral tolerance (Milich *et al.*, 1990; 1993). For this reason mutations which prevent the production of the HBeAg mark the end of the immune tolerance phase in infected individuals. Thus mutations in the core gene appear at, or shortly after, the appearance of the precore stop codon mutation, although they have also been found in wild type precore sequences. The precore stop mutation involves a G → A switch at nucleotide 1896. Since this nucleotide bonds with the thymine (T) at position 1858 during RNA folding, the mutation results in an additional AT bond which stabilises the 3D folding of the ϵ encapsidaton signal. Specimens which have a cytosine (C) at this position do not gain pgRNA stability from the mutation and as a result rarely harbour the stop codon mutation. Since a C at this position is definitive of genotype A and F the precore mutation is genotype dependent. In my database the genotype dependent nonsense mutation at codon 633 (Li *et al.*, 1993) is present in 18 of 65 specimens which included specimen E03 as well as a third of specimens from genotype D, 41.7 percent of genotype B specimens and 27.7 percent of genotype C specimens. The database has no mutations to the precore initiation codon (genotype independent) although two genotype A specimens have nonsense mutations at positions other than amino acid 633. Specimen A09 has a G → A switch at nucleotide 1897 and specimen A'02 has a deletion affecting the reading frame of the precore transcript at nucleotide 1929 and a nonsense mutation at codon 665. It was speculated that the G → T mutation at nucleotide 1862 could also prevent HBeAg expression (Kramvis *et al.*, 1997). This mutation is definitive for the common South African subtype A' but specimen A'01 has no non-definitive core gene variation. Figure 39 groups the 65 specimens according to the amount of individual core gene change from most

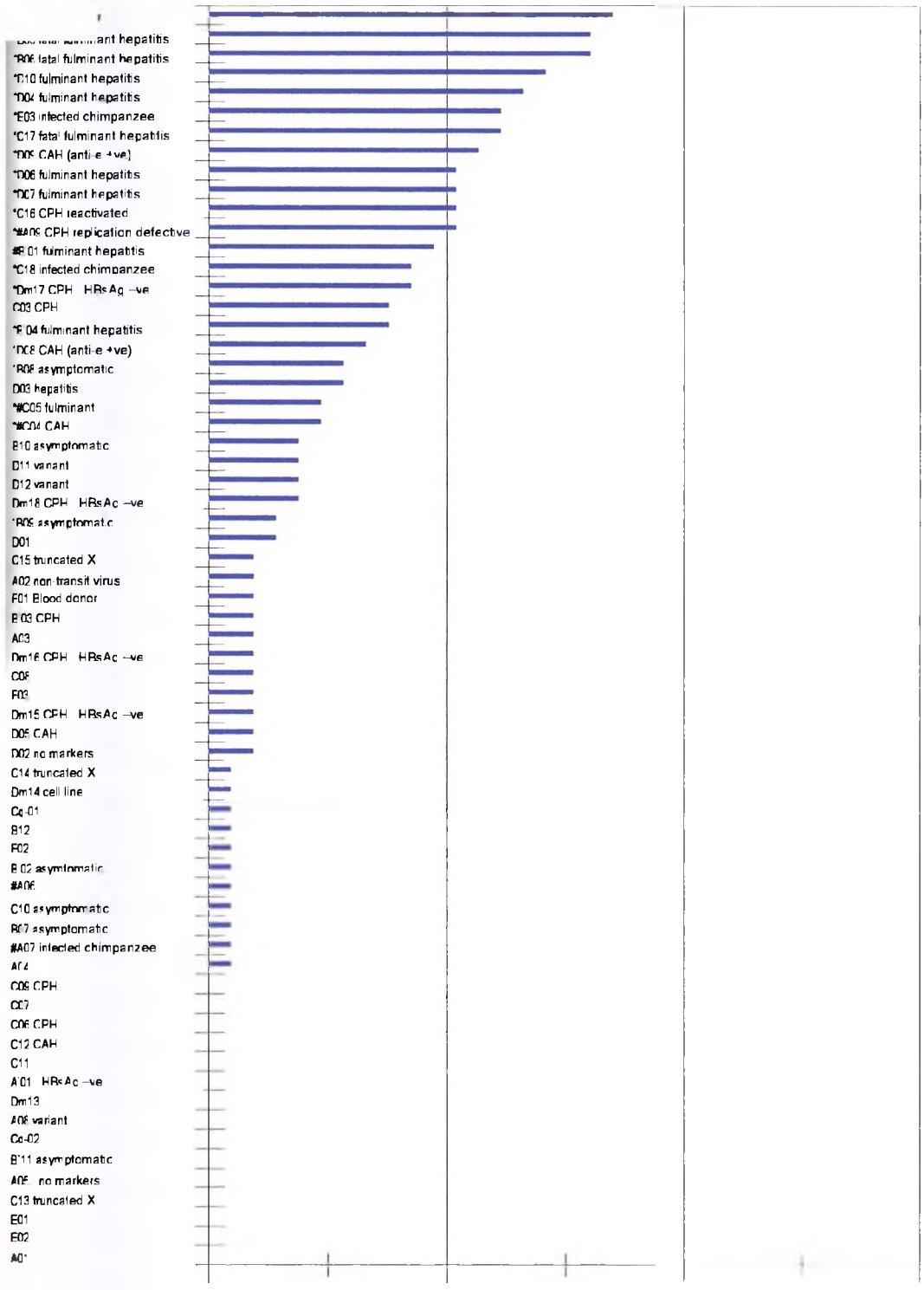


Figure 39. Non-definitive change per amino acid in each of 65 full-genome specimens in the precore/core domain

* indicates preCore stop mutants

# indicates core promoter mutants

to least and shows that all of the fulminant group of patients have considerable non-definitive change in the core gene and that 17 of the 22 specimens with the most change have the 1896 core gene mutation. With the exception of A'02 and A09 all genotype A, E and F specimens are in the lower half of the figure.

Various mutations clustering in the 25 bp upstream of the transcription initiation site of the precore transcript (1744-1751) reduce the efficiency of this transcript (Wasylyk *et al.*, 1981). Fifteen specimens (23 percent) from the database had mutations at either ^{1762}T or ^{1764}A, or both. The statistics of this and other GenBank databases probably do not reflect the prevalence of mutations in nature as abnormal specimens are more likely than typical specimens to be sequenced and studied. Baptista *et al.* (1999) reported that the combination of missense mutations (^{1762}T and ^{1764}A) had a prevalence of 66 percent in black Africans with HCC as compared with asymptomatic carriers where the prevalence was 11 percent. When non-definitive change was adjusted for the length of the domain, the greatest non-definitive change per amino acid was found to be in the core and RT domains with the least in the preS1 and preS2 domains.

Subtype variation at class I or class II anchor residues was found in 12 of 41 published epitopes (see Table 2 for full list of published epitopes). Some of these such as $^{43(RF1)}$Phe → Ser, $^{73(RF1)}$Lys → Ser, $^{94(RF1)}$Lys → Ile/Pro/Arg, $^{101(RF1)}$Leu → Pro, $^{245(RF1)}$Ser → Lys, $^{565(RF2)}$aspartic acid (Asp) → tryptophan (Try), $^{840(RF2)}$valine (Val) → cysteine (Cys)/Ser, $^{990(RF3)}$Asn → Ala/Arg and $^{1042(RF3)}$Ala → Ile/Asn/Asp are not conservative changes altering the properties and/or type of the protein (Taylor 1986) which could allow immune escape and minimize immune-mediated hepatitis. Others such as $^{50(RF1)}$Ala → Val, $^{261(RF1)}$Lys → Val, $^{661(RF1)}$Val → Ile, $^{750(RF1)}$Lys → Ile, $^{574(RF2)}$Val → Lys and $^{1035(RF3)}$Ile → Met/Val, $^{43(RF1)}$ are conservative changes and would not be expected to drastically affect MHC complex binding or TCR recognition of their respective epitopes. However an antagonistic effect caused by conservative amino acid substitutions has been reported (Bertoletti *et al.*, 1994) in viral strains infecting chronic HBV patients. Also when individual residues of the entire epitope were examined, many of the residues, and not just the anchor residues, were found to interact with the MHC or the TCR, sometimes having a dramatic effect on the induction of the CTL response even when the binding of defined synthetic peptides to purified MHC class I molecules was only decreased 10-fold (Bertoletti *et al.*, 1997).

Finally I examined sites reported to be important for transcription and replication. The glucocorticoid receptor was found to be mutated in three specimens (B"11, B"12 and C06), the enhancer I site at nucleotide 1183 was altered in four specimens (A01, B'03, B'04 and C12) and the enhancer II site (2726) defective in only one specimen (B'03). D11 was the only specimen with a modified TATATAA box in the promoter region of preS1, Dm17 had a G → C mutation just after this polyA site and four specimens were not conserved in the direct repeat (DR) regions. The three core promoter mutants (C13-C15) all reported to be serologically silent for HBV had an altered DR1 region while C02 was altered in the DR2 region. Associating pathogenesis with viral variation is no simple task but as data accumulates the changes which are important to the various disease states should become clearer.

CHAPTER 7

RELATIONSHIPS WITHIN AND BETWEEN GENOTYPES OF HEPATITIS B VIRUS AT POINTS ACROSS THE GENOME: FOOTPRINTS OF RECOMBINATION IN CERTAIN ISOLATES

7.1 Introduction

In highly endemic areas, HBV infection is predominantly acquired during the perinatal/ neonatal period or by horizontal transmission in the first few years of life (Beasley and Hwang, 1984; Botha *et al.*, 1984; Vardas *et al.*, 1999). Since this results in a high prevalence of long term HBV carriers with a low average age at infection (Edmunds *et al.*, 1996), the virus has a long time span in which to evolve within its host. A nucleotide exchange rate of between 0.1 and 0.7 per year (Günther *et al.*, 1999) has been estimated for the HBV (Okamoto *et al.*, 1987) and WHV (Girones and Miller, 1989) genomes respectively, which is similar to the most slowly evolving gene of retroviruses, the *gag* gene, and one to two orders of magnitude lower than the mutation rates previously calculated for the positive- and negative-strand RNA viruses (Girones and Miller, 1989).

Onginally, four genotypic groups of HBV (A to D) were defined, based on an intergenotypic divergence score of 8.5 to 10.0 percent between 18 complete genomes, as compared to a score of 1.1 to 2.7 percent between isolates within the same genotype (Okamoto *et al.*, 1988). This genotypic classification was extended to 6 genotypes (A to F) by phylogenetic analysis of 122 surface antigen (HBsAg) genes (Norder *et al.*, 1993). Norder *et al.* (1994) defined a subgroup of genotype C, Cq-, which lacked the *q* determinant and I have characterized two subgroups of genotype A, subgroups A and A' in an urban population from South Africa (Bowyer *et al.*, 1997). One of the aims of this study was to examine the relationships between full genomes of HBV in order to determine whether further subgroups to the major genotypes exist. In particular, I was interested in the relationship between genotypes D and E, since the existence of genotype E as a unique monophyletic group has been questioned (Kidd-Ljunggren *et al.*, 1995), and my screening assay has revealed that isolates from genotype E are common in the North Western parts of southern Africa.

Another important aim was to clarify and characterize the incongruence I (Bowyer *et al.*, 1997) and others (Mizokami *et al.*, 1997) have observed between trees constructed from different parts of the HBV genome. I speculated as to the possibility that recombination could be responsible as Georgi-Geisberger *et al.* (1992) had observed footprints of recombination in integrated HBV sequences. More recently Bollyky *et al.* (1996) presented statistical evidence that recombination was responsible for mosaic variation in HBV. At the same time I clarified the partitions, and sub-partitions, of HBV and derived a screening assay to identify mosaics within HBV isolates (Bowyer *et al.*, 2000). The anomalous fragments which I identified were then characterized with particular reference to their position in conserved and variant regions of the HBV genome.

7.2 Materials and Methods

7.2.1 Patients

The 65 patients were the same GenBank specimens used to map the "hotspots" and definitive sites in Chapter 6.

7.2.2 Phylogenetic analyses

Sequences were aligned manually using the DNASIS genetic systems software programme (Hitachi software). Multiple HBV DNA phylogenies and bootstrap analyses were performed using programmes (DNAPARS, SEQBOOT, DNADIST, FITCH, NEIGHBOR, DNAML, CONSENSE, DRAWTREE and DRAWGRAM) from the PHYLIP phylogeny inference package (version 3.5c, by Joseph Felsenstein). Sequences were identified by their GenBank accession numbers as well as by their position in the DNAPARS consensus tree (Figure 40 and paragraph 6.2.1). As before, subgroups were defined as groups of isolates divergent by 4 percent or less and clusters were combined to form higher order clades if their component isolates showed a sequence divergence of less than this. Retaining the standard cut-off of 8 percent divergence to define genotypes, the 65 specimens were classified into a series of 11 subtypes within the 6 genotypes.

7.2.3 Average genetic distance graphs

The aligned full-genome sequences were split into 53 separate files of 60 nucleotides and one of 42 nucleotides using CLUSTALW (Thompson *et al.*, 1994). These 54 files of inter-

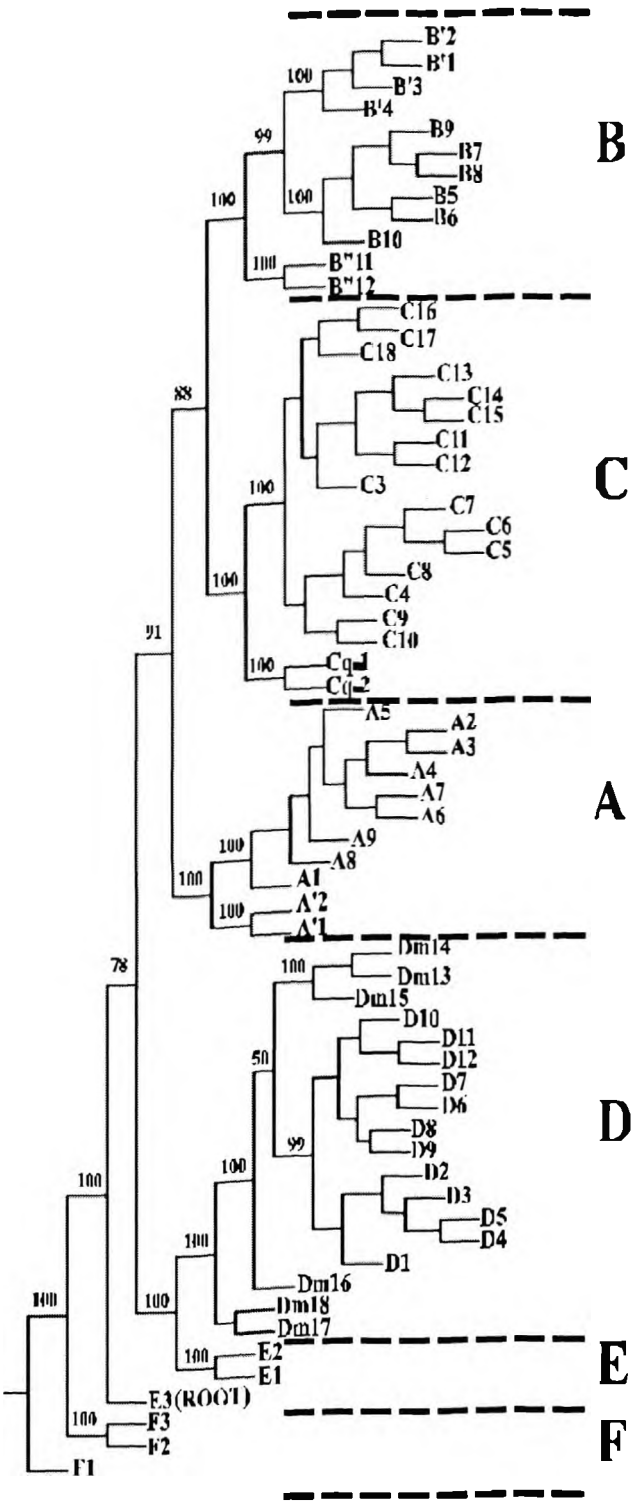


Figure 40. Full-genome FITCH bootstrapped consensus tree (100 sets) showing the conventional segregation of HBV specimens into the 6 genotypic groups (A-F). Bootstrap values for the most robust groupings are shown. Although all trees are essentially unrooted, D00220 was chosen as the outgroup and/or seed when the programme requested it

leaved sequence data provided the input to DNADIST which produced 54 distance matrices, one for each 60 nucleotides of the genome. Each matrix was formatted into a single column and imported into the standard spreadsheet QUATTRO PRO (Corel 8, Perfect Office Suite) for ease of subsequent calculation and manipulation. For each of the 11 subgroups at each of the 54 intervals, I then calculated the average pair-wise distance between:

1. isolates from the same subgroup
2. isolates from different subgroups
3. and isolates from subgroups belonging to different genotypes.

Plotting these average distances against the position of the nucleotide interval within the genome, generated 11 intra-subgroup, 6 inter-subgroup and 49 inter-genotype genetic distance graphs.

7.2.4 Isolate screening

Each of the 65 specimens was screened for conformance to type using a simple self-written dBase programme based on Siepel's recombinant identification program, RIP (Siepel and Korber, 1995). My programme compared the sequence of each specimen with each of 11 consensus sequences (one for each subgroup) and recorded the number of matches over each of 50 nucleotides. Variant regions within a specimen were identified when the best match within a window switched from the type established from the full-genome consensus tree.

7.2.5 Individual genetic distance graphs

Having identified specimens within the database which contained variant regions, I plotted, in turn on the same axis, the average pair-wise distance of the specimen from subgroups of interest at each of the 54 points along the genome.

7.2.6 Nucleotide/protein maps

The distribution and nature of nucleotide (and amino acid) mutations within an isolate was mapped against the consensus sequence of the parental subtypes. The base (or amino acid) at each variant position was listed together with the corresponding base/amino acid from its original (on the left) and alternate (on the right) genotype. If either reference base/amino acid matched the specimen it was replaced by an asterisk. Proteins were mapped in all three reading frames (not shown).

7.2.7 Subgenomic bootstrap trees

The boundaries of the variant regions defined 9 mosaic blocks, or fragments, within the specimens and the bootstrapped re-sampling NEIGHBOR-JOINING tree was drawn for each and compared with the full-genome bootstrapped tree.

7.2.8 Histograms

I examined the distribution of the genetic distances between the 65 specimens at conserved and variable domains along the genome. These included the preS1 region, the surface gene (which overlaps the RT/ *Pol* domain of P), the RNAse H domain, the X gene, the core gene, the TP domain and the preS1 region. The distance matrix for each region was used to calculate the frequency of each successive 0.005 range of genetic distance. This frequency was plotted to show the distribution of genetic distances within each of three categories: intra-subgroup, inter-subgroup and inter-genotype. Database records included information on the source of each distance reading so that a listing of specimens contributing to peaks of interest in the histograms could be generated, sorted and analyzed when required.

7.3 Results

There have been many genetic classification studies of HBV (Kidd-Ljunggren *et al.*, 1994; Kidd-Ljunggren *et al.*, 1995; Mizokami *et al.*, 1997; Ohba *et al.*, 1995; Norder *et al.*, 1993) since the first phylogenies appeared (Okamoto *et al.*, 1988; Uy *et al.*, 1992; Norder *et al.*, 1992). To clarify some of the questions raised by these and other studies, I divided the 6 genotypes into 11 subgroups in which specimens differed from each other by 4 percent or less and used sequence similarity measurements to examine the relationships between 65 complete genomes at 54 points along the genome.

7.3.1 Subgroups of HBV

Using the distance matrix programme FITCH, with 100 data sets, the 65 complete genomes grouped into the 6 conventional genotypes (A-F) all with bootstrap values of 100 percent (Figure 40). Genotype E isolates clustered together and away from genotype D. Genotype F was the most diverse group, separate from all other genotypes.

Regardless of the algorithm used, the bootstrapped trees all showed two clusters of genotype

A, three of genotype B, and five of genotype C. I retained the designations A and A' to differentiate the subgroups of genotype A (Bowyer *et al.*, 1997). Each of the three subgroups of genotype B cluster with an original prototype of genotype B as recognized by Okamoto *et al.* (1988) and were designated B, B' and B" (for the groups containing D00329, D00330 and D00331 respectively, previously designated serotype *adw1*, *adw2* and *adw3*. Despite numerous clades of C in the bootstrapped tree, only the subgroups C and Cq- observed by Norder *et al.* (1994) fitted my subgroup criteria. Genotype D forms a core clade with a very high bootstrap value of 99 plus 6 outliers which were designated Dm (for mutant group). I treated the three isolates of genotype F as a single clade despite the fact that the average difference between them (calculated from the full-genome matrix) was 4.7 percent, outside the subgroup range. Thus I defined a total of 11 subgroups (A, A', B, B', B", C, Cq-, D, Dm, E and F) of the 65 isolates within the 6 conventional HBV genotypes (A-F).

7.3.2 Average genetic distance graphs

Full-genome sequences were compared within and between subgroups at each of 54 positions along the genome.

7.3.2.1 Intra-subgroup

Figure 41(a) compares the average pair-wise distances between specimens within each of the 11 subgroups in my chosen range of 0 to 4 percent. Subgroup A of genotype A is the most conserved subtype showing intra-subgroup distances below 4 percent over most of the genome [Figure 41(a) i]. The effect of treating genotype F (within which specimens varied by 4.7 percent) as a single subgroup is evident in Figure 41(a) ix. The pattern of this genetic distance graph is more reminiscent of the inter-genotypic graphs [Figure 41(b)] confirming that more than one subgroup of this genotype exists and demonstrating the value of the graphs in providing a snapshot of the relationships between isolates from the same or different subgroups.

7.3.2.2 Inter-subgroup

The inter-subgroup graphs show a series of conserved and variable domains along the genome of HBV [Figure 41(b)]. Surface gene variation is typically below 4 percent between specimens in subgroups of the same genotype. The X gene is also well conserved and differences seldom exceed 8 percent. The preS1 and preS2 regions vary by 8 percent or more ex-

(a)

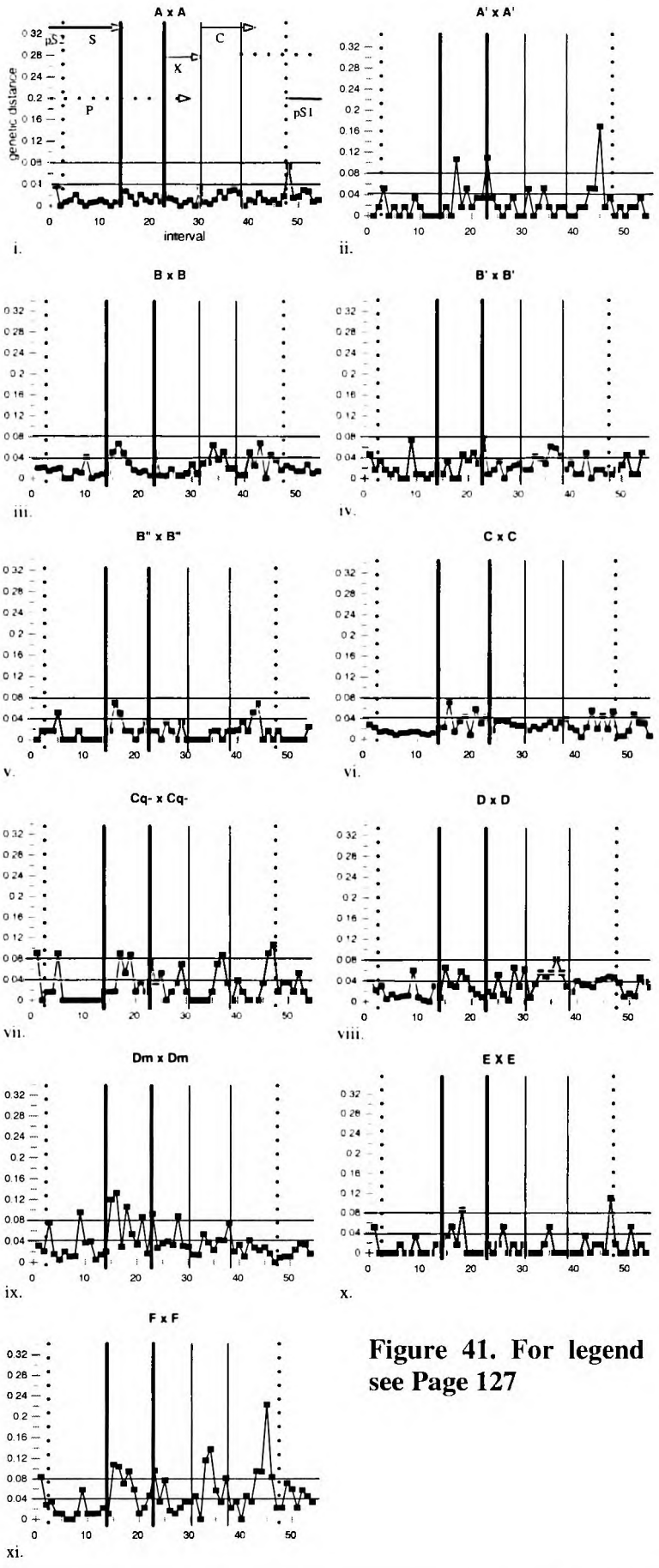


Figure 41. For legend see Page 127

cept in the B x B' [Figure 41(b) ii] and D x Dm [Figure 41(b) vi] graphs. The TP domain of the P gene (intervals 38-48) peaks above 8 percent except in the D x Dm graph [Figure 41 (b) vi]. The RNAse H domain (intervals 18-27 between the heavy vertical lines in Figure 41) of the P gene is most variable between subgroup B" and the other two subgroups of genotype B [Figure 41(b) iii and iv]. Differences between subtypes in the core gene (intervals 30-41 between the thin vertical lines in Figure 41) vary from less than 8 percent between subgroups B' and B" [Figure 41(b) iv] to greater than 18 percent between B and B' [Figure 41(b) ii] and B and B" [Figure 41(b) iii].

7.3.2.3 Inter-genotype

After careful examination of the 49 inter-genotypic graphs, I selected 8 typical, or anomalous, graphs to illustrate the features of this series [Figure 41(c)]. The conserved and variable domains are more defined in these graphs. Differences of 12 to 34 percent are typical between the genotypes in the preS1, preS2 and TP domains. The heterogeneity of the early RNAse H domain persists. The S and X genes, particularly in their early regions, are most conserved. Unconstrained regions of the HBV genome show the most variation. These include the last third of the *Pol*/RT gene, the first half of the RNAse H domain (before it overlaps the X gene) and almost the entire TP domain (the boundaries of the domains of P are as defined by Miller, 1988). The preS1 and preS2 regions which overlap the spacer region of the polymerase gene also fall into this category. Although the second half of the X gene and the first two-thirds of the core gene are unconstrained, the former is well conserved and the latter is not uniformly variable between the different types. The anomalous peaks in B x B' and B x B" [marked with * in Figure 41(b) ii and iii] are dramatically reversed in their corresponding inter-genotype graphs with genotype C [again marked * for B' x C {Figure 41(c) iv}, and for B" x C {Figure 41(c) v}] These two sets of graphs show quite clearly that two entire subgroups of genotype B are more closely related to genotype C over at least 480 nucleotides (intervals 30-37). In contrast, the graph of subgroup B (the third clade of genotype B) versus subgroup C [Figure 41(c) iii] has the typical inter-genotype pattern. Observe the inter-subgroup differences between both subgroups of genotype D and genotype E over the latter part of the X gene and most of the C gene [Figure 41(c) vi and vii].

(b)

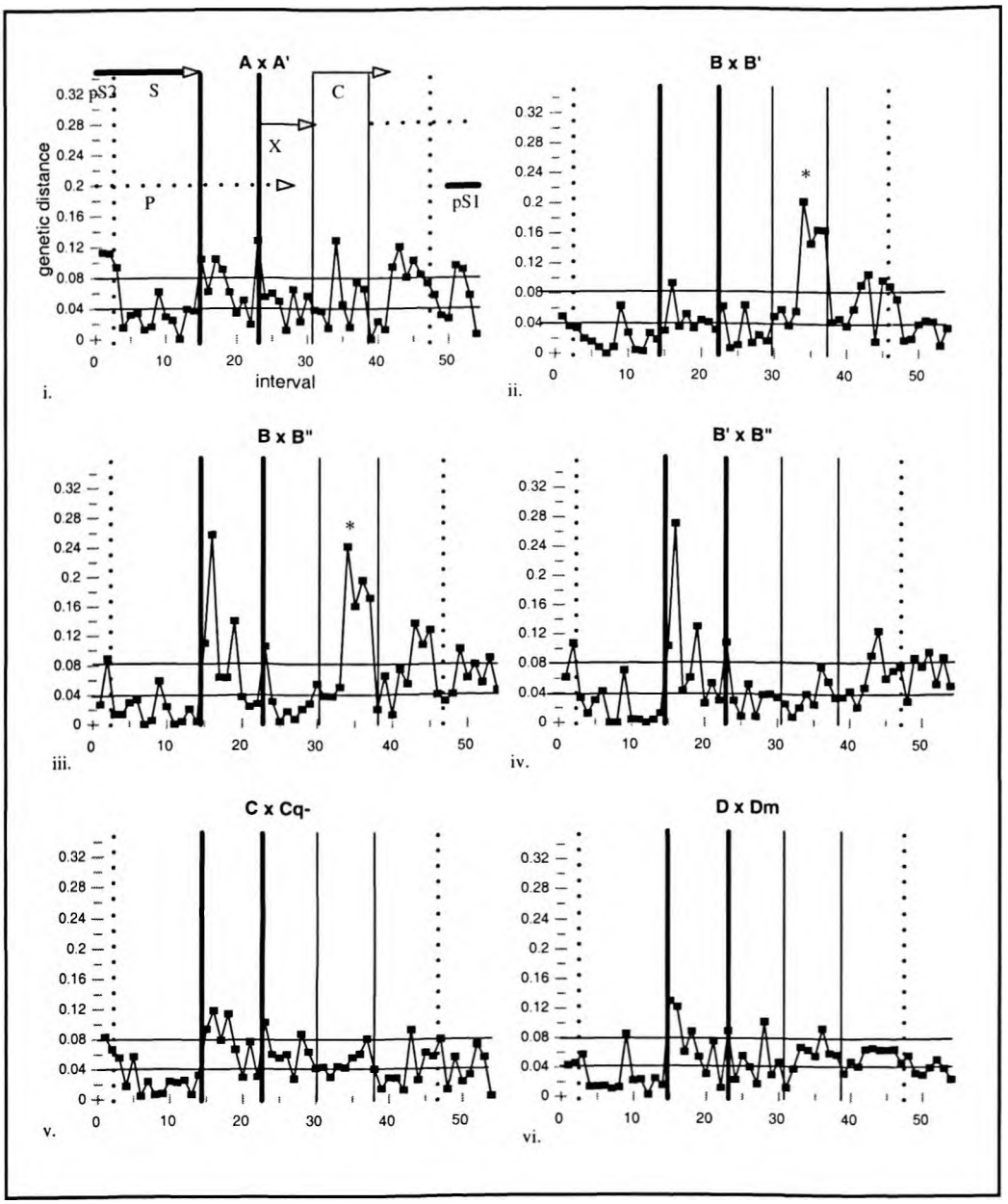


Figure 41. For legend see Page 127

7.3.3 Mosaic structure within isolates

The screening programme identified mosaic sequence in 14/65 isolates. These isolates were from subgroup D (3/12, D05-07), subgroup Dm (3/6, Dm16-18), subgroup B' (4/4, B'01-04), subgroup B" (2/2, B"11-12) and genotype E (2/2, E01-02). In all cases genotype D contained mosaics of genotype A, and genotype B contained mosaics of genotype C. Both genotype E specimens displayed only subgroup differences from the genotype D subgroups between nucleotide 1576 and 2262.

(c)

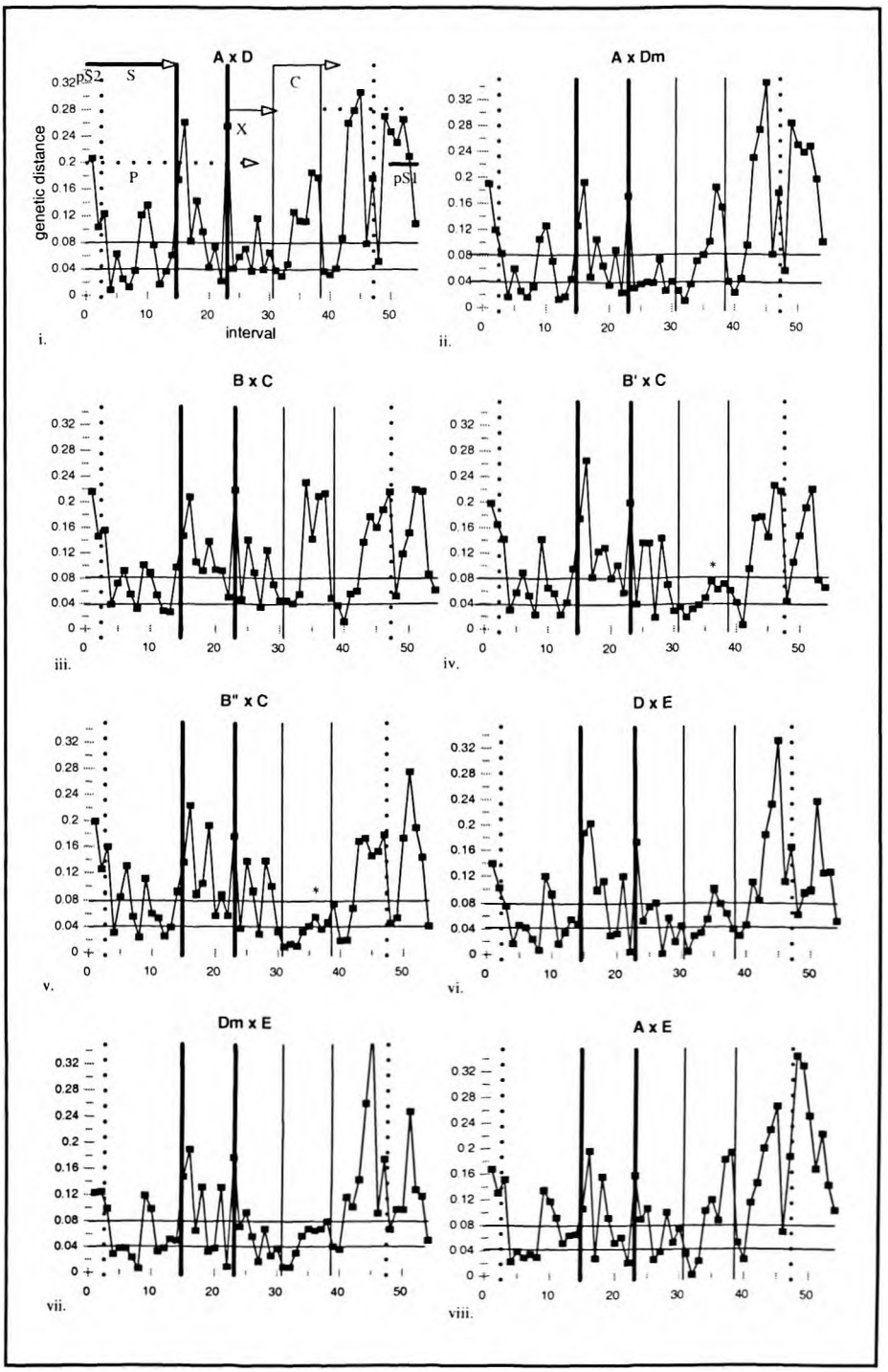


Figure 41. Interval number (on the x axis) is plotted against average genetic distance: (a) within subgroups, (b) between subgroups of the same genotype, and (c) between selected subgroups of different genotypes. The asterisks in Figure 41(b) draw attention to the anomalous subregions within the B' x B and B" x B inter-subgroup graphs and in Figure 41(c) to the corresponding regions within the B' x C and B" x C inter-genotype graphs. Vertical lines demarcate the beginning and end of genes preS, S, P, X and C of HBV, in the 3 reading frames [see Figure 41(a) i], and horizontal lines highlight the 4 and 8 percent group category cutoffs. Zero is the *Eco R1* site and each interval consists of 60 nucleotides

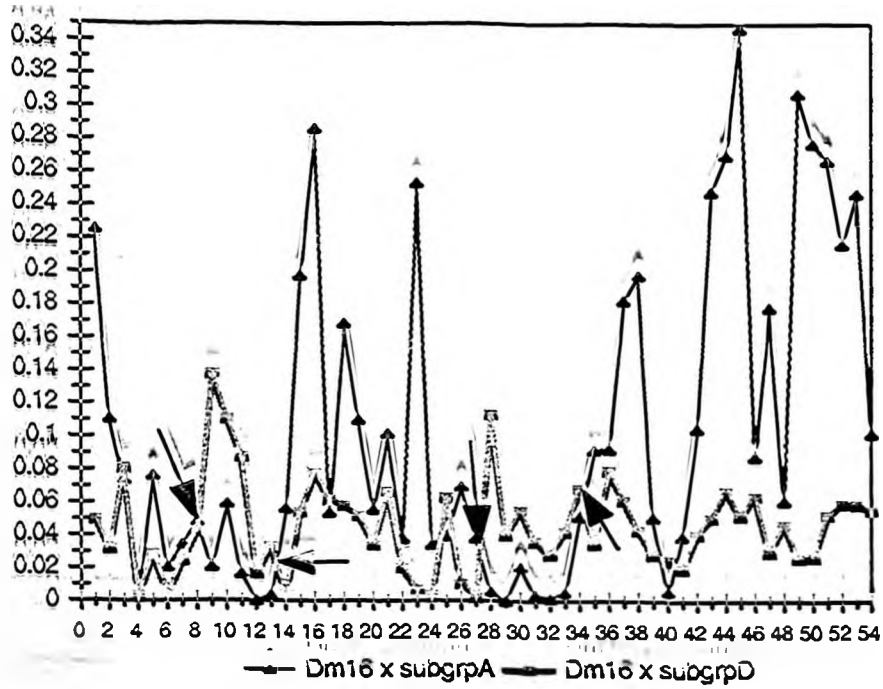
Graphs of genetic distance of variant specimens from the consensus of their parental genotypes located breakpoints. Figure 42 shows representative graphs from specimens Dm16 [Figure 42 (a)] and B'01 [Figure 42(b)], respectively. The first and last variant nucleotide [Figure 43(a)] and amino acid (not shown) of each block was then precisely identified. Nine mosaic fragments were identified, within the 14 specimens [Figure 43(b)].

Bootstrap values (with a 75 percent cut-off) were used to give a statistical measure to the genotype switch within the fragments and to establish the confidence of the groupings (Simmonds *et al.*, 1996). Mosaic structure was demonstrated in 7 out of 9 fragments [Table 15(a) bootstrap column]. The genotype D specimens Dm16 [Figure 42(a), 43(b) and 44(a)], Dm17 [Figure 43(a) and (b), and 44(b)] and Dm18 [Figure 43(b), 44(b) and (c)] have one, or more, extensive mosaic segments which cluster with genotype A in more than 78 percent of the trees. A 459 bp fragment in the core gene region [Figure 42(b), 43(b), 44(c) and (d)] which clusters with genotype C 100 percent of the time was common to all specimens in the two genotype B clades B' and B". Two of the shorter fragments, fragment I in specimen Dm18 and fragment VII present in specimens D05-D07, which clustered with genotype A did so in less than 75 percent of the trees and therefore were too short to be classified as recombinants with statistical significance (Table 15). There is a complete disruption of the conventional phylogeny of HBV in the core sub-genomic region [Figs. 44(c) and (d)]. In this region, genotype B is represented by subgroup B specimens only while the rest of genotype B clusters with genotype C and genotype E loses its own identity and clusters with genotype D. Genotype A loses its identity as a separate and discrete group between intervals 28 and 30 unless variants (D05-D07 and Dm16-Dm18) are excluded from the analysis (not shown).

7.3.4 Effect of parallel evolution

If selection is causing parallel replacements, the evidence for mosaic structure should be stronger when only synonymous change is considered (Smith, 1992). However, the opposite was true within 8 of the 9 fragments. The difference between a fragment and its respective parental consensus sequences when total change is considered [abbreviated mosaic index or MI in Table 15(a)] is greater than when synonymous change alone is considered [synonymous mosaic index or SMI in Table 15(b)] with one exception, Fragment V in RF1 [see MI-SMI < 1, Table 15(b)]. Thus mosaic structure does not appear to have evolved through parallel evolution.

(a)



(b)

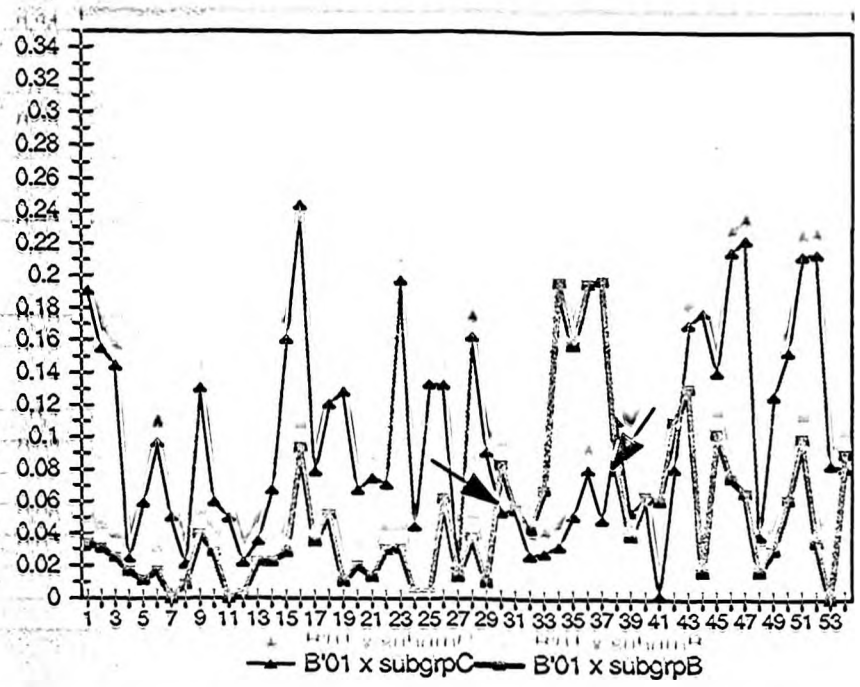


Figure 42. Graphs of pair-wise distance of putative recombinants from parental subgroups to illustrate breakpoint determination. Average pair-wise distance over each of the 54 intervals between (a) isolate Dm16 and i. subgroup A (black with triangular markers) and ii. subgroup D (thick grey with square markers) and (b) isolate B'01 and i. subgroup C (black with triangular markers) and ii. subgroup B (thick grey with square markers). The graphs in Figure 42(a) intersect at four places within intervals 8 and 13 and also within intervals 27 and 34 indicating two separate inserted fragments in specimen Dm16. The graphs in Figure 42(b) intersect only within intervals 30 and 37 (see black arrows) indicating a single inserted fragment in this specimen

Table 15. Similarity of fragments derived from mosaic areas to parental subgroups

Frag No	Specimen	G1	G2	Position		Intervals		length	G1(nt)	% G1(nt)	G2(nt)	% G2(nt)	Bootstrap %	Mosaic Index
				start	stop	start	stop							MI=%G2(nt)-%G1(nt)
I	Dm18	D	A	129	402	3	7	274	12	95.6	6	97.8	NS	2.19
II	Dm16	D	A	495	780	9	13	286	19	93.4	4	98.6	93	5.24
III	Dm18	D	A	737	1219	13	21	483	43	91.1	8	98.3	100	7.25
IV	Dm17	D	A	822	1775	15	30	954	82	91.4	20	97.9	100	6.50
V	Dm18	D	A	1326	2082	23	35	757	40	94.7	13	98.3	92	3.57
VI	E01	A	D	1576	2339	27	39	764	55	92.8	18	97.6	95	4.84
VII	D5	D	A	1637	1775	28	33	139	10	92.8	3	97.8	NS	5.04
	D6	D	A	1637	1759			123	9	92.7	2	98.4	NS	5.69
	D7	D	A	1637	1759			123	9	92.7	2	98.4	NS	5.69
VIII	B1	B	C	1742	2200	30	37	459	46	90.0	14	96.9	100	6.97
	B2	B	C	1742	2200			459	41	91.1	14	96.9	100	5.88
	B3	B	C	1742	2200			459	36	92.2	14	96.9	100	4.79
	B4	B	C	1742	2200			459	47	89.8	16	96.5	100	6.75
	B*11	B	C	1742	2200			459	45	90.2	9	98.0	100	7.84
	B*12	B	C	1742	2200			459	45	90.2	6	98.7	100	8.50
IX	Dm18	A	D	2191	2337	38	39	147	15	89.8	1	99.3	78	9.52

Frag No	Specimen	G1	G2	Position		RF	gene	length	G1(aa)	G2(aa)	SG1	SG2	% SG1	% SG2	Syn. MI			
				start	stop										SMI=%SG2-%SG1	MI-SMI	SG1/G1	SG2/G2
I	Dm18	D	A	43	134	1	S	92	6	2	6	4	97.8	98.54	0.73	1.46	0.50	0.33
				43	134	3	P	92	3	2	9	4	96.7	98.54	1.82	0.36	0.75	0.33
II	Dm16	D	A	165	260	1	S	96	9	2	10	2	96.5	99.30	2.80	2.45	0.53	0.11
				165	260	3	P	96	7	3	12	1	95.8	99.65	3.85	1.40	0.63	0.05
III	Dm18	D	A	246	278	1	S	33	3	0	40	8	91.7	98.34	6.63	0.62	0.93	0.19
				246	407	3	P	162	13	0	30	8	93.8	98.34	4.55	2.69	0.70	0.19
IV	Dm17	D	A	459	592	2	X	134	7	1	75	19	92.1	98.01	5.87	0.63	0.91	0.23
				274	540	3	P	267	21	2	61	18	93.6	98.11	4.51	1.99	0.74	0.22
V	Dm18	D	A	606	684	1	C	85	0	1	40	12	94.7	98.41	3.70	(0.13)	1.00	0.30
				459	612	2	X	154	7	1	33	12	95.6	98.41	2.77	0.79	0.83	0.30
VI	E01	A	D	606	780	1	C	175	7	3	48	15	93.7	98.04	4.32	0.52	0.87	0.27
				525	612	2	X	88	3	1	52	17	93.2	97.77	4.58	0.26	0.95	0.31
				525	540	3	P	16	0	0	55	18	92.8	97.64	4.84	0.00	1.00	0.33
VII	D5	D	A	547	592	2	X	46	6	1	4	2	97.1	98.56	1.44	3.60	0.40	0.20
	D6	D	A	547	592	2	X	46	5	1	4	1	96.7	99.19	2.44	3.25	0.44	0.11
	D7	D	A	547	592	2	X	46	5	1	4	1	96.7	99.19	2.44	3.25	0.44	0.11
VIII	B1	B	C	606	734	1	C	129	9	2	37	12	91.9	97.39	5.45	1.53	0.80	0.26
	B2	B	C	606	734	1	C	129	10	1	31	13	93.2	97.17	3.92	1.96	0.76	0.32
	B3	B	C	606	734	1	C	129	10	2	26	12	94.3	97.39	3.05	1.74	0.72	0.33
	B4	B	C	606	734	1	C	129	10	2	37	14	91.9	96.95	5.01	1.74	0.79	0.30
	B*11	B	C	606	734	1	C	129	11	0	34	9	92.6	98.04	5.45	2.40	0.76	0.20
	B*12	B	C	606	734	1	C	129	11	0	34	6	92.6	98.69	6.10	2.40	0.76	0.13
IX	Dm18	A	D	730	779	1	C	50	5	0	10	1	93.2	99.32	6.12	3.40	0.67	0.07

The mosaic areas identified in Figure 43(b) were divided into nine fragments (I-IX) and each fragment was characterised by a bootstrap tree. In (a) and (b), G1 and G2 refer to the original and alternate genotype of the specimen, respectively. G1 (nt) [or G1 (aa)] refers to the number of nucleotides (or amino acids) within the fragment which differ from the original parent; G2 (nt) [or G2 (aa)] refers to the number of nucletides (or amino acids) within the fragment which differ from the alternate genotype.(a) The similarity of each fragment within each specimen to both parental subgroups was determined as a percentage [columns '% G1 (nt)' and '% G2 (nt)', respectively)] The difference in similarity between the parental subgroups [Mosaic Index (MI) = % G2 (nt) - % G1 (nt)] is listed. NS, not significant. (b) The extent of each fragment in each reading frame [RF column) was determined at the amino acid level and again the similarity of each specimen to both parental subgroups was determined. The latter was used to determine the synonymous change (SG) in each fragment of each specimen in each reading frame and the difference in synonymous change between the parental subgroups [synonymous mosaic index (SMI) = %SG2—%SG1] is shown. A positive value in column 16, MI-SMI, indicates that the mosaic character is less enhanced when synonymous change alone is considered (and vice versa). The ratio of synonymous change versus total change for each specimen in each reading frame from original genotype and alternate genotype are given in columns SG1/G1 and SG2/G2, respectively

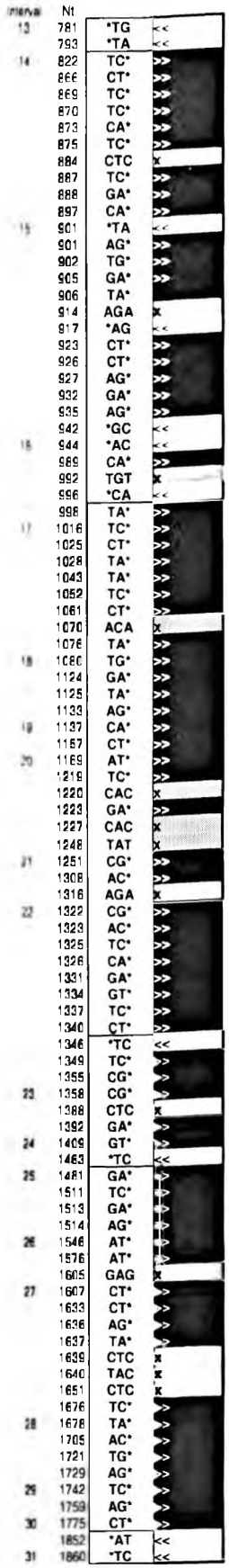


Figure 43(a) For legend see Page 133

7.3.5 Effect of host selective pressure

I examined type/subtype variation at anchor residues of known linear epitopes (Chisari and Ferrari, 1995) relative to the 9 fragments and their variant amino acids. Subtype variation at anchor residues was found in 17/31 epitopes examined. Nine of these were represented in the fragments: preS2-44-53 (Phe/Ser), HBs-185-194 (Ser/Ala), Pol-816-824 (Asp/Val), Core-1-20 (Ser/Thr), Core-18-27 (Val/Ile), Core-28-47 (Asp/glutamic acid [Glu]), Core-50-69 (Asn/Thr), Core-88-96 (Thr/Val) and Core-111-125 (Leu/Ile).

7.3.6 Effect of functional constraint

The ratio of synonymous change to total change is a measure of functional constraint within a gene or the degree to which a gene is conserved. A functionally important gene will be well conserved with a ratio close to one since in the main change is synonymous. Conversely, genes ordered according to their increasing substitution rate are also ordered for decreasing functional importance. The ratios for change from original genotype [Table 15(b), SG1/G1 column] show a difference in the functional constraint between the different reading frames, genes and positions on the genome while the ratios for change from alternate genotype are fairly constant [Table 15(b), SG2/G2 column].

7.3.7 Histograms

The 8 frequency histograms which plot the distribution of genetic distances at specific regions of the genome (Figure 45) give an indication of the rate at which the genotypes are changing in relation to one another. The genetic distances were plotted in three separate series (intra-subgroup [black], inter-subgroup [white] and inter-genotype [grey]) and these formed

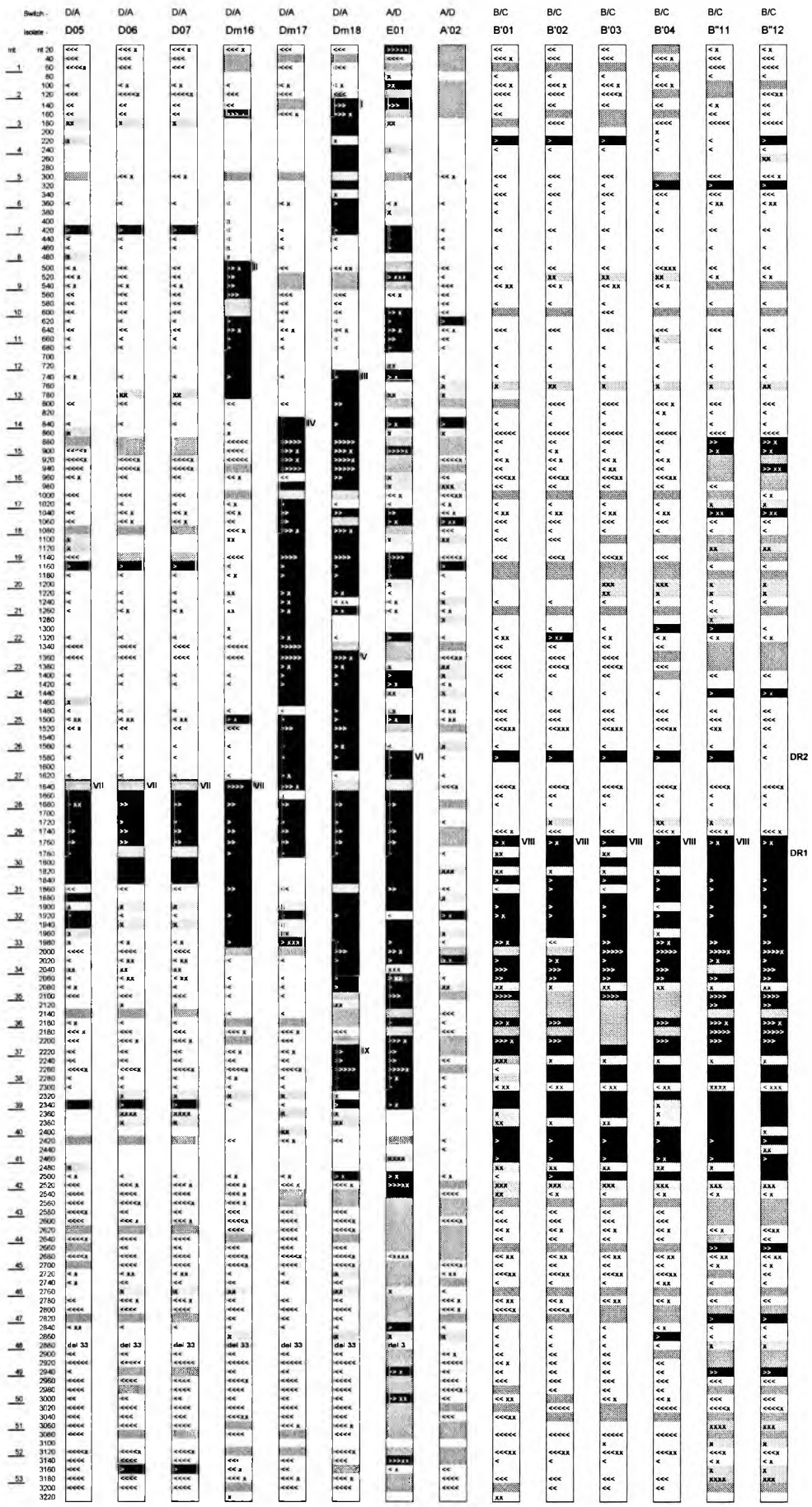


Figure 43(b) For legend see Page 133

three distinct but overlapping distributions in most of the histograms. The median and range of the isolate and subtype distributions do not vary greatly across the genome with the notable exception of the sub-genomic fragment corresponding to the surface gene where the subtype/isolate distinction is minimal [Figure 45(a) ii]. On the other hand, the range and median of the type distribution varied greatly across the genome. Multiple distributions and a wide range of genetic distance were present in the preS1 [Figure 45(a) i] and preS2 [Figure 45(a) vii] sub-genomic regions. Examination of the specimens contributing to these multiple distributions confirmed that the genotypes are not evolving at a constant rate over this part of the genome. The genotypic median does not exceed 8 percent in the surface [5.8 percent, Figure 45(a) ii] and X [7.9 percent, Figure 45(a) iv] gene regions but ranges from 8.9 percent to 14.9 percent over the other sub-genomic fragments.

Specimens containing mosaic blocks are "wrongly" grouped in variant sub-genomic regions so their presence is detected by inter-subgroup distances in the inter-genotype range (or vice versa) and appear as a right-hand bias of the white inter-subgroup distribution (or left-hand bias of the grey inter-genotype distribution). This effect is most prominent in the core gene sub-genomic fragment [Figure 45(a) v]. The right-hand bias of the grey inter-genotype distribution, present in many of the histograms, was formed primarily by the pair-wise distances between specimens from genotypes A-E and specimens from genotype F. This is evident in all the histograms but is most marked in the X region [Figure 45(a) iv] and the TP domain [Figure 45(a) vi] and forms a discrete distribution in the full-genome histogram [(Figure 45(b)].

Figure 43. The distribution and nature of variation within individual isolates mapped against the consensus sequence of the parental subtypes as described in the text. (a) Mosaic region (nucleotide 822-1775) of specimen Dm17. For clarity only variant sites are listed. Over these 954 nucleotides the fragment would require 82 changes [69 A/D (black, right arrowheads) and 13 unique/quasispecies changes (grey, x)] to conform with the consensus of the original genotype D of Dm16 but only 20 changes [7 D/A (white, left arrowheads)] and the 13 unique/quasispecies changes] to conform with the alternate genotype A. (b) By coding each batch of 20 nucleotides in the same way as before, the complete genome, and not just variant positions, of 13/14 putative recombinants and one non-recombinant, A'02, is mapped proportionately to show mosaic areas in context. The black areas mark the position and extent of the mosaic regions within each specimen. Roman numerals I-IX show the start of each of the nine mosaic fragments characterised and listed in Table 15

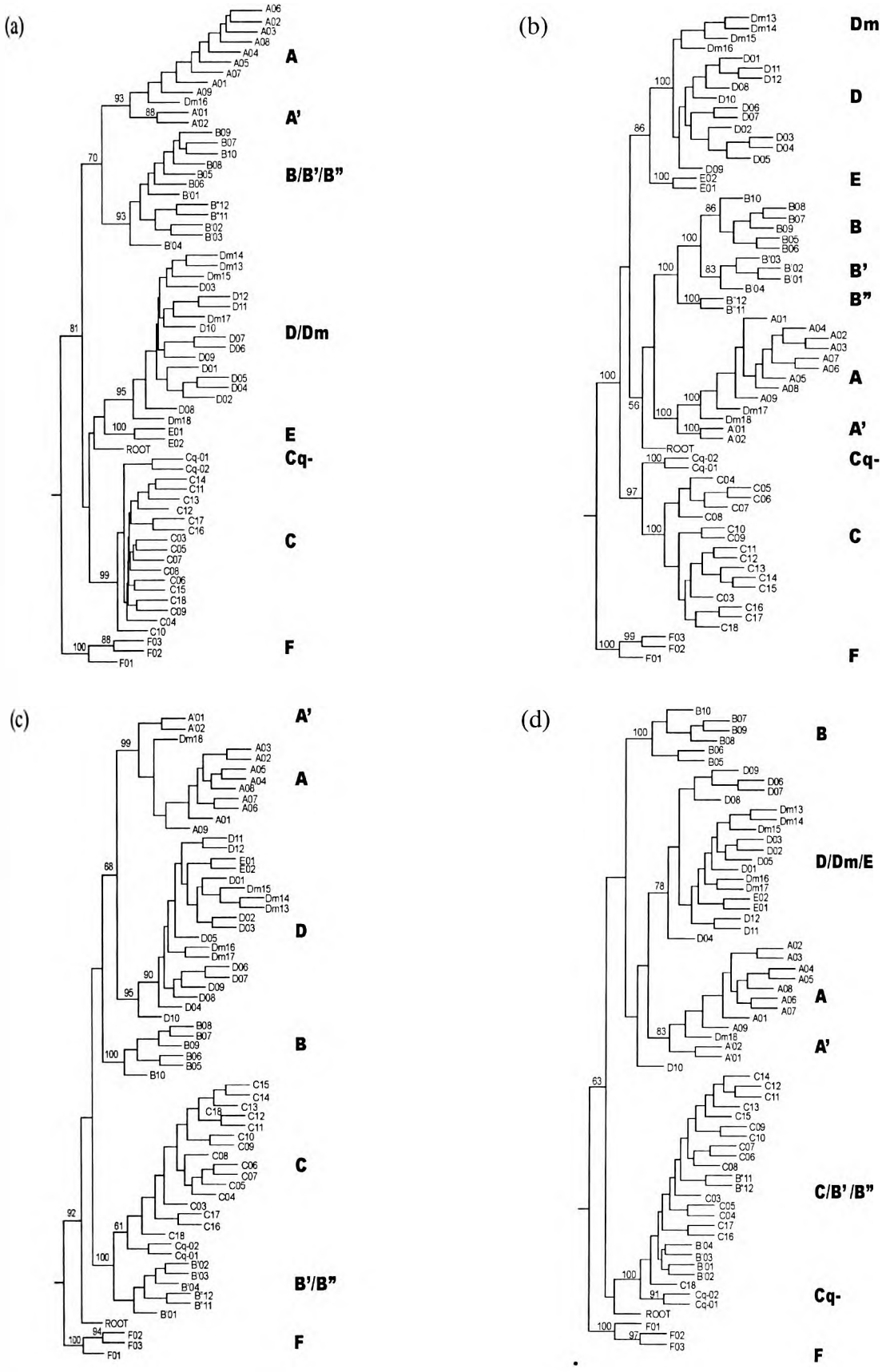


Figure 44. For legend see Page 135

7.4 Discussion

Type, subtype and isolate categories of HBV were determined from the full-genome bootstrapped tree (Figure 40). I plotted average evolutionary distances within and between categories at each of 54 points along the genome (Figure 41) and determined the distribution of evolutionary distances within sub-genomic domains of HBV (Figure 45). The genetic distance graphs showed a distinct pattern of alternating variable and conserved domains within and between HBV subtypes. Genotype B does not have the same topology over different sub-genomic regions. Two of the subgroups of B, B' and B", cluster with genotype C in the core gene region [Figure 44(c) and (d)]. B and B' and D and Dm did not appear to be true partitions and the lack of preS variation in their intra-subgroup graphs appeared to confirm this [Figure 41(b) ii and vi]. Genotype E did not partition as a separate monophyletic group over all sub-genomic regions, often clustering with genotype D [44(c) and (d)]. Genotype F is extremely different from the other genotypes and, although some variable sites are shared between genotypes B and F, none of the three inter-genotypic graphs between the clades of genotype B and genotype F, or the trees, showed a recent relationship between these two genotypes. Only genotypes A (subtypes A and A'), C (subtypes C and Cq') and F (not defined) partition consistently into subtypes.

Figure 44. Bootstap resampling trees of 4/9 mosaic fragments. (a) Fragment II, intervals 9-13, present in Dm16 only. Dm16 clusters with genotype A with a bootstrap value of 93 percent. Note that subgroups of genotypes B and D are not present over this subgenomic fragment whereas the subgroups of genotype A, A and A', and genotype C, C and Cq-, are. Genotype E clusters as a separate group with a bootstrap value of 100 percent. (b) Fragment IV, intervals 15-30, is present in Dm17 only but includes parts of fragments III and V of Dm18. Both Dm17 and Dm18 cluster with genotype A with a bootstrap value of 100 percent. Note the subgroups of genotypes A, B and C are present over this subgenomic fragment whereas genotype D shows Dm13 to Dm16 as outliers of the main cluster. Genotype E clusters as a separate group with a bootstrap value of 100 percent. (c) Fragment VI, intervals 27-39, present in E01 and E02. Genotype E clusters with genotype D with a bootstrap value of 95 percent. The subgroups of genotypes A and C are visible over this subgenomic fragment but 2 of the 3 subgroups of genotype B cluster with genotype C. (d) Fragment VIII, intervals 30-37, present in B'01 to B'04 and both B"11 and B"12 also includes part of Fragment V of Dm18. All the group B specimens cluster with genotype C with a bootstrap value of 100 percent and the subfragment of Dm18 clusters with genotype A with a bootstrap value of 83 percent. Note the subgroups of genotypes A and C are present over this subgenomic fragment and that Genotype E clusters with genotype D with a bootstrap value of 78 percent

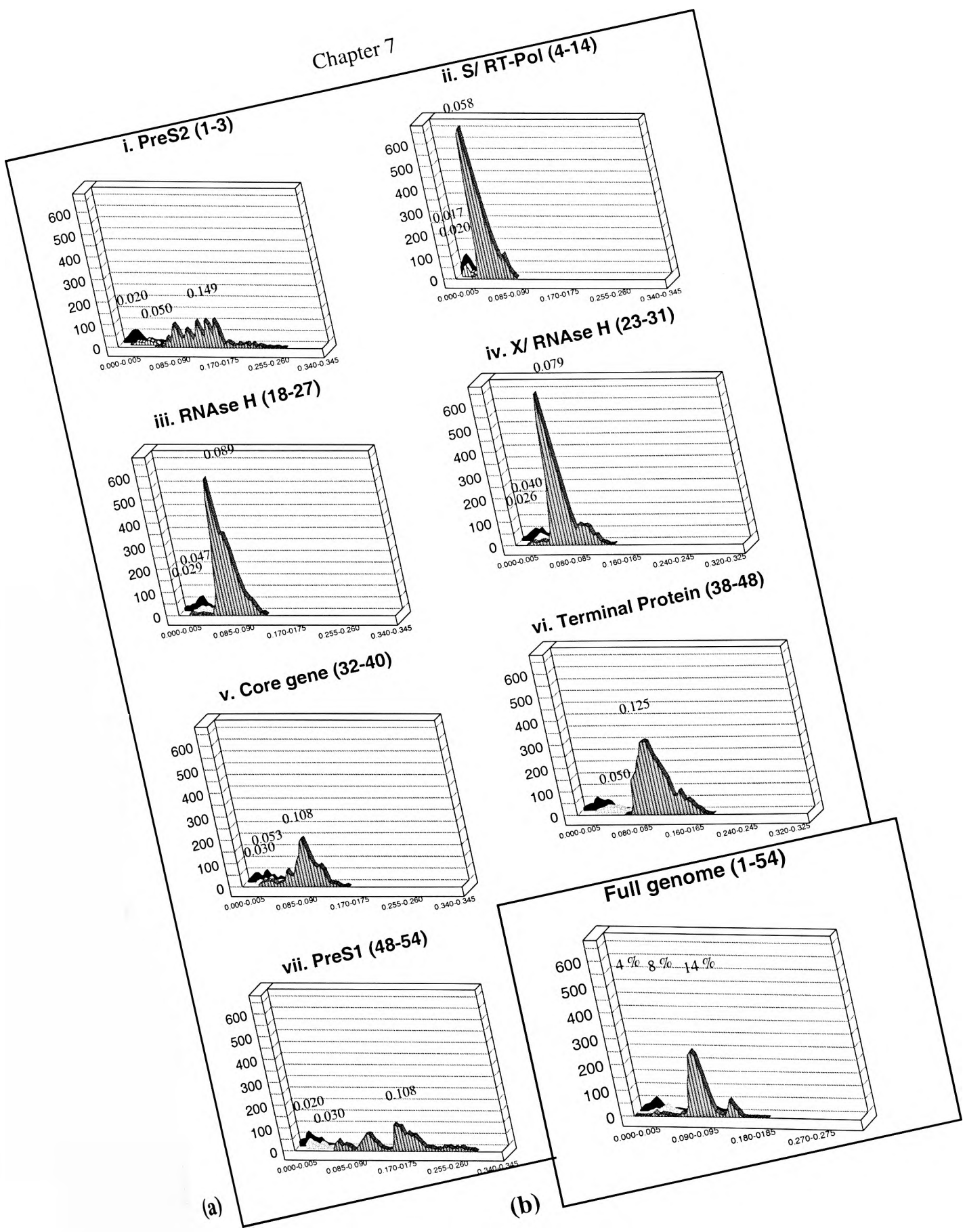


Figure 45. For legend see Page 137

The histograms showing the distribution of pair-wise distances (Figure 45) over sub-genomic regions indicate that substitution rates are not constant among the lineages in the preS regions and between genotype F and all other genotypes, A to E, over most of the genome. This is contrary to the findings of Yang *et al* (1995a), but genotypes E and F and subgroups B, B' and Cq were not represented in their study.

Screening of isolates individually, 50 nucleotides at a time, against the consensus sequence of the 11 subgroup partitions identified possible mosaic structure in 14 specimens. Mosaic structure was demonstrated statistically in 11 isolates using bootstrap resampling (Table 15). The nature and distribution of polymorphic sites within the fragments was mapped at both the nucleotide (Figure 43) and protein levels (not shown). The position of the fragments was related to mutational "hotspots" and linear epitopes of HBV. Fragments were found in all except the preS1 coding region. The sequence between and including the two DR regions (nucleotides 1592-1840) is represented in all recombinant specimens. A common region, fragment VIII, was involved in all subtype B' and B" specimens. This same fragment was found in four additional subgroup B' isolates from GenBank (X98073 - X98076) which were excluded from the main study because their genomes had large insertions and/or deletions.

Sequence variation may be due to chance point mutations or recombination of DNA segments (Stephens, 1985). At the same time, functional constraints, host-viral interactions and selective pressures determine the mutations which are lost and those which are retained and this can lead to gene conservation or parallel evolution in independently arising strains. Frequent recombination and/or mutational "hotspots" can confuse evolutionary relationships, but in the simplest case, mosaic blocks of sequence identical to an alternate type (or subtype) within a specimen of established type is considered unequivocal evidence that recombination has taken place (Smith, 1992). Mosaic structure caused by random parallel replace-

---

Figure 45. The distribution of pair-wise genetic distances (a) over each of 7 named sub-genomic genes/domains, the intervals covered by each region are shown in parentheses, or (b) over the complete genome. Successive ranges of genetic distance along the X axis progress in steps of 0.005, while the Y axis represents the frequency of that range within the distance matrix for the region being studied. The contribution of each of the three distributions is separate: intra-subgroup distances (black); inter-subgroup distances (white); inter-genotype distances (grey)

ments would be more evident when only synonymous change is considered but this was not found within the identified fragments. Functional constraints limit non-synonymous variation and result in variable subtype/type differences in HBV at the protein level (Mizokami *et al.*, 1997). This was evident when the translation products of the fragments were compared with the consensus proteins of their original genotype. However, when compared with the consensus of their alternate genotype, a constant difference was observed. This would only be expected when like proteins of the same subtype are compared and further supports the mosaic structure within the fragments. Some of the changes to the linear epitopes observed within the fragments e.g. preS2-44-53 (Phe → Ser) would be expected to alter the binding characteristics of host HLA antigens while many e.g. Val → Ile, Thr → Val, Leu → Ile or Ser → Ala would not be expected to cause a major change. Although the ends of Fragments V and VI correspond approximately to the end of the major epitopes of the core gene, many epitopes of P (Pol-61-69) and all preS1 epitopes are not represented in the fragments. Thus different major histocompatibility complex (MHC) backgrounds in human populations in different parts of the world are unlikely to be entirely responsible for the heterogeneity I have observed. Bootstrap resampling confirmed the mosaic structure in 11 specimens and recombination, rather than random change, appears to be the dominant mechanism for this structure.

None of the mosaics observed in this study breach the known geographic boundaries of the genotypes, as established by molecular epidemiological studies (Magnius *et al.*, 1995). This is a necessary condition of my mechanism of choice since recombination implies a relatively high frequency of superinfection. Superinfection has been reported for HBV but this has always been considered rare and unimportant (Heijtink *et al.*, 1982; Tabor *et al.*, 1977). HBV replication involves template switches during both minus (Wang and Seeger, 1993; Tavis *et al.*, 1994) and plus (Will *et al.*, 1987) strand synthesis and intra- or inter-molecular template switching is a common mechanism for homologous recombination (Pathnak and Wei-Shau, 1997). However, this is thought to be an unlikely mechanism in HBV since the HBV pregenome replicates only after encapsidation (Ganem, 1991) and, unlike the retroviruses which have a dimeric genome, HBV is thought to package a single RNA pregenome. Nevertheless, Raimondo *et al.* (1988) have reported the presence of replicative intermediates sensitive to deoxyribonuclease I (DNAse I) digestion in the liver of a patient, and suggested that unencapsidated molecular forms of HBV DNA can accumulate in chronic HBV carriers. Chronic

camage has a complex pathology progressing from replicative to non-replicative disease and often resulting in viral integration and/or hepatocellular carcinoma. This progression is not always linear and non-replicative carriers can re-activate and return to an earlier stage of the disease (Dusheiko *et al.*, 1985). Further studies are needed to confirm that template switching is impossible at all stages of disease progression.

However, replication is not the only stage at which HBV has an opportunity to recombine. Initiation of infection and hepadnavirus replication involve conversion of genomic relaxed circular DNA (RC DNA) into covalently-closed circular DNA (cccDNA) within the nucleus of the infected cell in a manner not fully understood but which is thought to utilize cellular DNA modifying enzymes (Kock and Schlicht, 1993). It has also been speculated that cellular enzymes could be responsible for changes to the episome making it a better substrate for integration (Schirmacher *et al.*, 1995). Increasingly it is being suggested that the processes of mutation and integration are linked in some instances. Identical mutations have been reported in free and integrated WHV (Kew *et al.*, 1993) and HBV from a single patient (Georgi-Geisberger *et al.*, 1992). A recent study used an *in vitro* DHBV system to map the plus and minus strand cleavage sites of topoisomerase I (top I) which is considered a likely candidate for both conversion of relaxed circles to cccDNA and for integration of episomes into the host DNA (Pourquier *et al.*, 1999). This model showed that top I was capable of converting RC DNA to cccDNA *in vitro* and that this was achieved via non-homologous recombination. An earlier study which defined illegitimate replication of linear DHBV DNA in primary hepatocyte cultures also found that the 3' end of the minus strand efficiently participates in intra- and intermolecular non-homologous recombination to produce monomeric cccDNA or oligomeric forms in which monomers are joined near the ends in random orientation (Yang and Summers, 1995b). Although these oligomeric forms have not been found in viral particles nor shown to take part in illegitimate replication, they would have a similar mosaic structure to that which I have observed.

Recombination has been documented previously in HBV. A 196 bp region in the precore/core was found to enhance recombination *in vitro* in the presence of extracts from actively dividing cells (Hino *et al.*, 1991). Georgi-Geisberger *et al.* (1992) found evidence of homologous recombination, very similar to what I have found at the population level, when studying integrated and episomal HBV from a single patient. Bollyky *et al.* (1996) use boot-

strapped maximum likelihood trees and a randomization test to demonstrate mosaic structure statistically in 2/25 complete genome sequences and conclude that the heterogeneity which they observe is the result of recombination between viruses of different genomic and antigenic types.

Further study is required to produce direct evidence for recombination in HBV and to clarify the role of recombination in the heterogeneity of HBV. The mosaic structure which I and others have observed affects entire clades and alters the phylogeny of HBV over extensive sub-genomic fragments. Many enigmas of hepatitis B persist, including geographical differences in the pathology and evolution of HBV, and the variation in host immune responses to HBV in different individuals (Foster and Thomas, 1993). Recombination may present one mechanism that could be responsible for these differences.

CHAPTER 8

HLA FREQUENCY AND THE SUSCEPTIBILITY TO OR PROTECTION FROM HBV IN SOUTHERN AFRICA

8.1 Introduction

Archeological evidence indicates that the San and the Khoi occupied South Africa as early as 8,000 BC and that at this time their territory extended as far north as the Great Rift Valley. The San, or Bushmen, were hunters and gatherers, while the Khoi, or Hottentots, were nomadic cattle herders. Although the Khoi are slightly taller, the two groups are much alike in appearance and are often referred to as the Khoisan.

Language family groups are useful for tracing the history of people before recorded history, and Khoisan is one of four great language family groups which follow ancient migratory patterns of influence on the African continent. When the Khoisan occupied the southern half of Africa, the Afro-Asiatic Linguistic group occupied the North and East of the continent, following the Saharan routes of trade and travel, while the western bulge of Africa (south of the Sahara) was populated by the Niger-Congo Linguistic group and sandwiched between these two was the Nilo-Saharan group. During the last three millennia BC, however, the pattern changed and there was a great expansion of the Niger-Congo language into the "fly" belt (Oliver, 1991). This migration, called the Bantu migration, was one of the largest in human history. Figure 46 shows the distribution of languages and language groups in present day Africa (http://www.lib.ohio-state.edu).

There were three major waves of migration of the Niger-Congo speaking people. Each of these moved along several separate routes to southern Africa and brought with it the Bantu culture and tradition of horticulture and cattle breeding. The first wave of expansion created North Bantu while the second and third expansions developed into Western and Eastern Bantu. Thus early in their history, the Eastern and Western Bantu split into two major linguistic branches. The large Eastern African migrant group migrated between the Great Rift Valley and the Indian ocean coast through present-day Zimbabwe (Shona) and

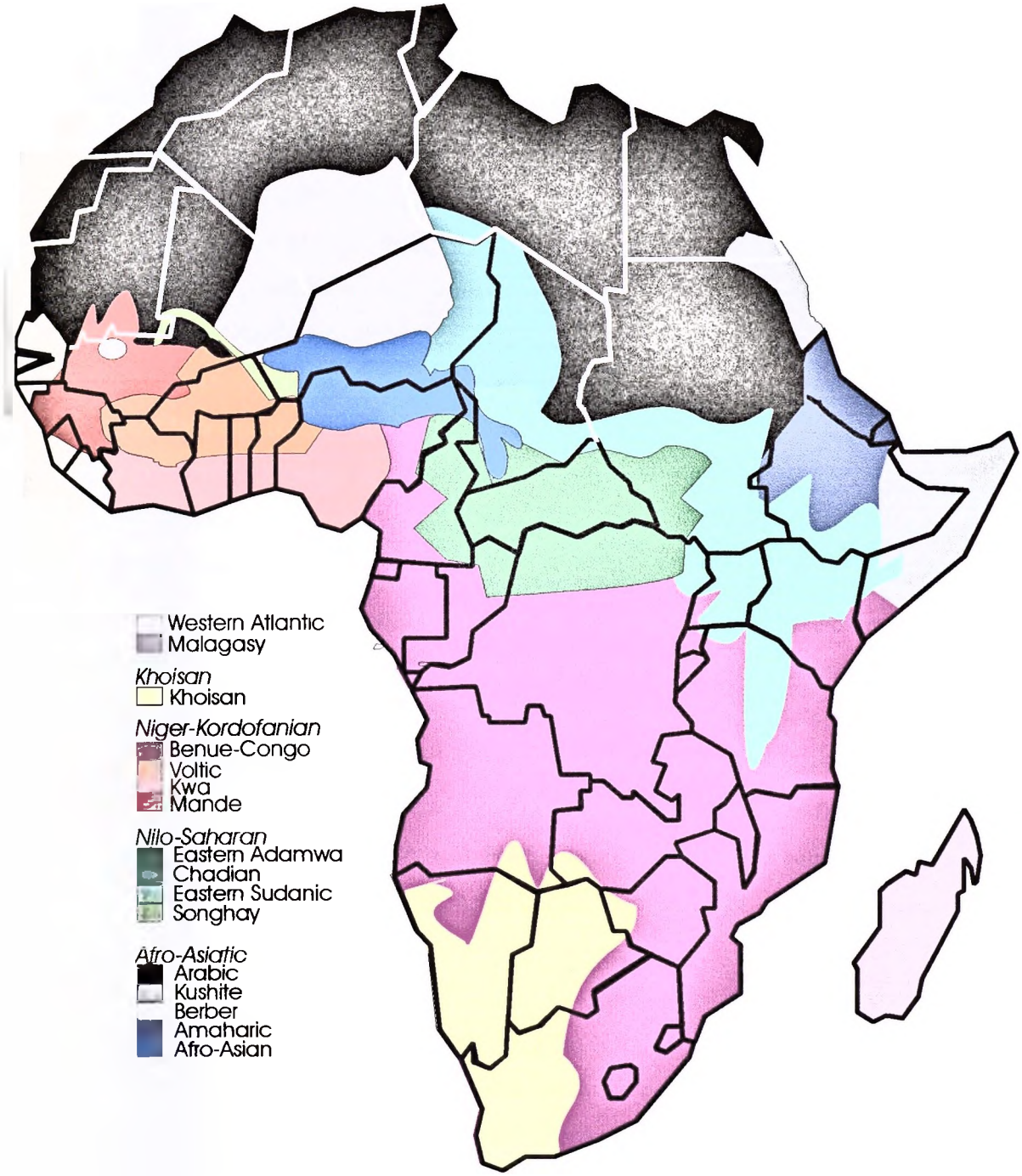


Figure 46. The distribution of languages from the basic African Linguistic groups in Africa today (adapted from map at the web site http://www.lib.ohio-state.edu)

Mozambique (Tsonga) down to South Africa. The Tsonga occupy southern Mozambique and adjacent areas in South Africa and encompass three sub-groups: the Ronga, Tswa and Tsonga (or Shangaan). The southern migrant group of the Eastern migration are known as the Nguni people. They followed an inland course via the headwaters of the most northerly bend of the Limpopo River bringing with them cultural attributes of West, Central and North-East Africa and in due course made contact with the San hunters. The "click" sounds which are characteristic of the Nguni languages today are a result of this interaction. The Nguni later split into separate migrations, moving in different stages into what is now KwaZulu-Natal and the Eastern Province. The Northern Nguni settled in northern Kwa-Zulu Natal while a portion of the people doubled back into what is now Swaziland and still others, the forebears of the Pondo, occupied the former Transkei. The Southern Nguni were the last to leave the Limpopo and settled for a while in the south-eastern region of the Mpumalanga province, then moved in stages into central KwaZulu-Natal. Finding the North-East and North-West already occupied, two smaller groups moved on. The first of these groups, the Xhosa, skirted around the Pondo, who occupied the coastal regions of the south, to settle the inland high ground while the second group, the Thembu, became the coastal neighbours of the Pondo. The final Nguni migration populated the heart of KwaZulu-Natal succeeding the Ndwandwe and Mthetwa empires respectively in the North-West and North-East establishing the Zulu kingdom. The latter, under the famous Zulu chief Shaka, conquered and displaced dozens of nearby small kingdoms to become the rulers of KwaZulu-Natal from the Tugela River in the South to the border of Mozambique in the North.

The Western Bantu moved into what is now Angola, Namibia (Ovambo) and north-western Botswana (Sotho-Tswana) moving into the land inhabited by the San Bushmen. The sons of the first great Sotho chief, Masilo, were the ancestors of the three main Tswana tribes. The Sotho were still moving slowly south in the 1800's eventually occupying the entire central plateau almost to the Orange River as well as the land which is now Lesotho. However, they were repeatedly invaded by marauding Nguni groups which pushed the Tswana (or Ntwana) westward across the Vaal River. One such group, the Ndebele of Zulu origin who were fleeing Shaka and led by Mzilikazi, caused devastation as they moved through Tswana territory driving the Pedi and the Venda to the Northern edge of the central plateau and the Tswana chiefdoms to the fringes of the Kalahari. Eventually the Ndebele moved to what is

now Matebeleland in Zimbabwe, although the group of Ndebele still living in the Northern Province came to be Sotho speakers (Oliver, 1991; http://www.africasouth.co.za; http:// encarta.msn.com; http://www.ukans.edu). The links which exist between the tribes of the Western migration can be seen in several remnants of their shared culture. For example, the Ntwana (Tswana) on the interior plateau of South Africa and the Ovambos in the North-West of Namibia both use horns for adornment and share the strange and unique helmet hair style which is a status symbol for the Ntwana (Tyrrell, 1996). The Ovambos make up about half of the people of Namibia. Together with the Kavango and the people of the East Caprivi, they occupy the northern part of the country and until recently had very little contact with the Nama, Damara and Herero who roamed the central part of the country. Today the San are found only in the north-western Cape, the Kalahari and parts of Namibia and Botswana. Both the Western Cape and central and southern Namibia have a relatively small black population but a large mixed, or coloured, population. They form an anthropologically distinct population of relatively recent origin of western European, southern African and south-east Asian descent. The southern African component was derived from the Khoisan with a small contribution from Madagascan slaves (Hammond *et al.*, 1997).

Apart from the four main African language families, two other language families now exist in Africa: The Indo-European languages introduced by recent colonists from South-East Asia and Europe and the Malayo-Polynesian languages, introduced by colonists from South-East Asia, which are exclusive to Madagascar. South Africa has 11 official languages of which 9 are African languages (Tshivenda, Sesotho, Sepedi, isiNdebele, Setswana, siSwati, isiXhosa, isiZulu and XiTsonga) evolved from the Benue-Congo languages of the Niger-Congo family. Figure 47 shows the distribution of the major ethnic groups in southern Africa at the present time.

This study examines the frequency of the major HLA types among the ethnic groups in southern Africa today and compares these with the HLA frequencies found within some of the major global population groups. Local HLA type was then related to known HBV disease associations. Finally a small group of local HBV acute and chronic patients and vaccinees were HLA typed and their phenotype compared with published population frequencies.

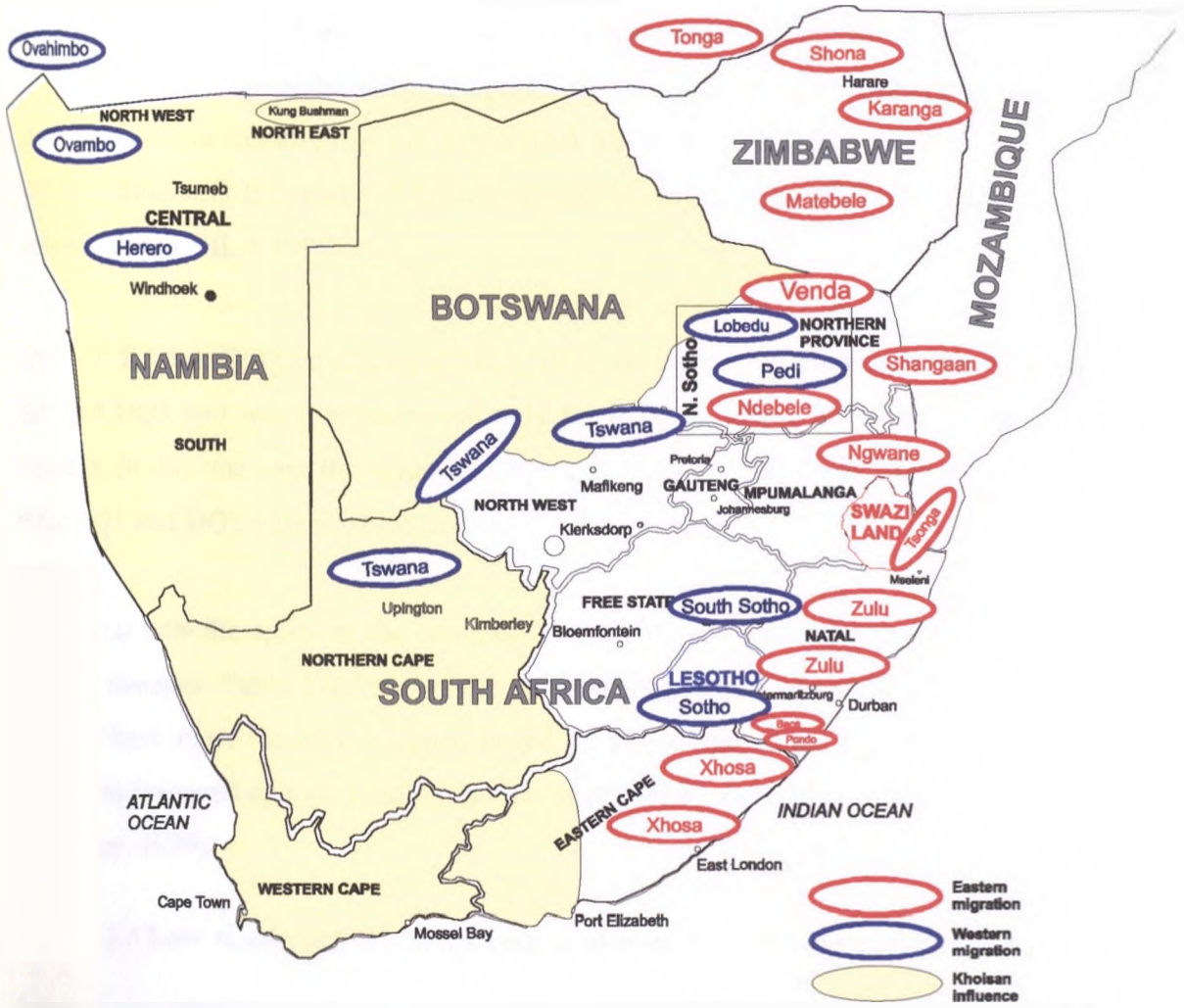


Figure 47. Distribution of major ethnic groups in southern Africa today (adapted from Tyrrell, 1996)

8.2 Materials and Methods

8.2.1 Patients

Two vaccinated South African Caucasians, four South African Blacks (3 males and one female) who were diagnosed with acute hepatitis B in Johannesburg hospitals and 7 chronic carriers from the North-East province of Namibia, all of whom had seroconverted to anti-HBe, were selected for HLA-DR typing. The diagnosis of acute hepatitis was based on the detection of IgM anti-HBc antibodies in the serum, the recent onset of jaundice and other typical symptoms of acute hepatitis.

8.2.2 Populations

Published HLA frequencies for 31 populations were compared at the Class I HLA-A and HLA-B loci. Twenty populations were compared at the Class II HLA-DR locus and 17 at

the HLA-DQ locus. A short description is given for some of the African populations studied (Table 16) together with the source of population data. Reference numbers refer to reference codes from the proceedings of the 11th (HLA 1991) or 12th (HLA 1997 accessed at http:// histo.chu-stlouis.fr) International Histocompatibility workshops (IHWC) or the 1998 update from this group (HLA 1998).

The Class II HLA-DQ frequencies in HLA 1991 did not narrow down the broad specificity of DQ1 and DQ3 and was therefore used only for the San, where no other source of data was available. In this one case the values of DQ5 and DQ6 and DQ8 and DQ9 were taken as half of the DQ1 and DQ3 values, respectively.

Table 17(a) lists the splits of the common major antigens while (b) lists the DR haplotypes and their lineages. Table 17(c) shows the nine supertypes into which the class I HLA-A and –B alleles have been recently assigned based on phenotype of the binding groove, and their subsequent favoured epitope binding motifs, of the MHC molecule for which they code (Sette and Sidney, 1999).

8.2.3 Low resolution HLA-DR typing of southern African patients

Patients were typed for HLA-DR using low-resolution sequence specific PCR (SSP) HLA typing kits from Dynal (Norway) as described in paragraph 3.3.5. I used high resolution-SSP HLA typing kits to confirm the typing of two specimens which were found to have rare alleles (DRB1*08 and DRB1*16). I also did high resolution typing on all specimens found to have the more common alleles DRB1*03 and DRB1*11.

8.3 Results

8.3.1 Class I HLA-A comparison

Figure 48 shows the HLA-A frequencies in selected populations colour coded where possible according to the four known supertypes of the HLA-A alleles (A2 [black], A 3 [yellow], A1 [blue] and A24 [red]; Sette and Sidney, 1999).

Allele A2 is represented in all the populations studied ranging from 7.8 percent in the Zulu to 27.8 percent in the Rimaibe. A28 which contains the splits A68 and A69 cannot be placed in a supertype since A*6802 and A*6901 are A2-supertype and A*6801 is A3-supertype. The

Table 16. A short description of some African populations studied and codes of ethnic groups with which they were compared

Ethnic Group	Ref. and/or code	Geographic region occupied	Language or Linguistic division	Comments
Amerinds	HLA 1991, 40800; 98-226			
Australian Aborigines	HLA 1991, 20100; 97-303			
British	HLA 1991, 30106; 97-208			
Bubi	Hammond *et al.*, 1997	Equatorial Guinea	North Bantu, Niger-Congo phylum	Inhabitants of Bioko island
Cape coloureds	du Toit *et al.*, 1988	Western Cape	Mixed population	Anthropologically distinct population of recent origin.
Fula	Allsopp *et al.*, 1992	Nomadic pastoralists of West Africa	West Atlantic	Subgroup of the Fulani, East African origin.
Indian	HLA 1991, 30117; 98-196			
Inuits	HLA 1991, 41102			
Italians	HLA 1991, 30117; 97-233			
Japanese	HLA 1991, 40101; 98-214			
Javanese	HLA 1991, 40414			
Jola	Allsopp *et al.*, 1992	Senegambia		
Khoikhoi	HLA 1991, 10202	Namibia	Khoisan	Also known as the Hottentots, only a few scattered groups remain
Mandinka	Allsopp *et al.*, 1992	Senegal & Gambia	Mande, Niger-Congo phylum	Originated in Mali expanded to West Africa 14th century
Manjago	Allsopp *et al.*, 1992			
Maoris	HLA 1991, 40602; 97-395			
Mossi	Allsopp *et al.*, 1992	Burkina Faso	Votaic, Niger-Congo phylum	Sedentary farmers of sub-Saharan savannah
Ndebele	HLA 1991, 97-179		isiNdebele	remnants of Kingdom established by Mzilikazi who fled the Zulu
Nigerians	Okoye *et al.*,1985	Nigeria	Bantu	
North African blacks	HLA 1991, 10100			
Rimaibe	Allsopp *et al.*, 1992	Burkina Faso		Closest ethnically to the Mossi
SA Caucasoids	du Toit *et al.*,1988			Descendents of Dutch, German, French & British.
Senegalese	HLA 1991, 10102; 97-177		West Atlantic	
Serere	Allsopp *et al.*, 1992			
Shona	du Toit *et al.*, 1988; 97-180	North Zimbabwe	Central Bantu	Descendents of the Karanga people
Sotho	97-181	Lesotho & Qwa Qwa in the Free State	Sesotho	Sesotho language 6.9 % of SA population
The San	du Toit *et al.*, 1990; HLA 1991, 10201	North Eastern Namibia & Botswana	Khoisan	
Vietnamese	HLA 1991, 40403; 97-307			
Wolofs	Allsopp *et al.*, 1992	North Gambia		Forced southward by Berber expansion in 13th century
Xhosa	du Toit *et al.*, 1988	Eastern Cape	IsiXhosa, South Eastern Bantu, Cape or South Nguni	17. 5 % of SA population, greatest contact with the Khoisan
Zaireans	HLA 1991, 10206; 97-176			
Zulu	97-182		IsiZulu, North or Natal Nguni, Central Bantu	22.4 % of SA population

All population HLA frequency data was taken from HLA 1991, HLA 1998, the web site http://histo.chu-stlouis.fr or the publications listed in column 2. Where the population used was referred to by a code in the reference this is also given

Table 17 (a). The major splits of the HLA-A, -B and –DR antigens (b) the five haplotypes and 13 lineages of HLA-DR and (c) alleles grouped into the 9 recognised Supertypes

(a)

Split Antigens of HLA-A	Split Antigens of HLA-B	Split Antigens of HLA-DR
A1	B5 B51, B52	DR1
A2	B7	DR2 DR15, DR16
A3	B8	DR3 DR17, DR18
A9 A23, A24	B12 B44, B45	DR4
A10 A25, A26, A34, A66	B13	DR5 DR11, DR12
A11	B14 B64, B65	DR6 DR13, DR14
A19 A29, A30, A31, A32, A33, A74	B15 B62, B63, B75, B76, B77	DR7
A28 A68, A69	B16 B38, B39	DR8
A36	B17 B57, B58	DR9
A43	B18	DR10
	B21 B49, B50	
	B22 B54, B55, B56	
	B27	
	B35	
	B37	
	B40 B60, B61	
	B41	
	B42	
	B46	
	B47	
	B48	
	B53	
	B59	
	B67	
	B70 B71, B72	
	B73	

(b)

DR Haplotype	DR lineage
DR1	DR1, DR10
DR8	DR8
DR51	DR15, DR16
DR52	DR3, DR11 and DR12 (DR5), DR13 and DR14 (DR6)
DR53	DR4, DR7, DR9

(c)

Supertype A2	Supertype A3	Supertype A1	Supertype A24	Supertype B7	Supertype B44	Supertype B27	Supertype 58	Supertype 62
A*0201	A*0301	A*0101	A*2301	B*0702	B18	B*1401-2	B*1516-17	B*1301-2
A*0202	A*1101	A*0102	A24	B*0703 5	B*3701	B*1503	B*5701-2	B*1501 (B62)
A*0203	A*3101	A*2501	A30	B*1508	B*4001	B*1509	B58	B*1502 (B75)
A*0204	A*3301	A*2601		B*3501-3	B*4006 (B61)	B*1510		B*1506
A*0205	A*6801	A*2602		B*5101	B*4101 (B60)	B*1518		B*1512
A*0206		A*2604		B*5301	B*4402	B*2701 8		B*1513 (B77)
A*6807		A*3201		B*5401	B*4403	B*3801-2		B*1514
A*6802		A*3601		B*5501-2	B*4501	B*3901-4		B*1519
A*6821		A*4301		B*5601-2	B*4901	B*4801-2		B*1521
		A*8001		B*6701	B*5001	B*7301		B*4601
				B*7801				B*52

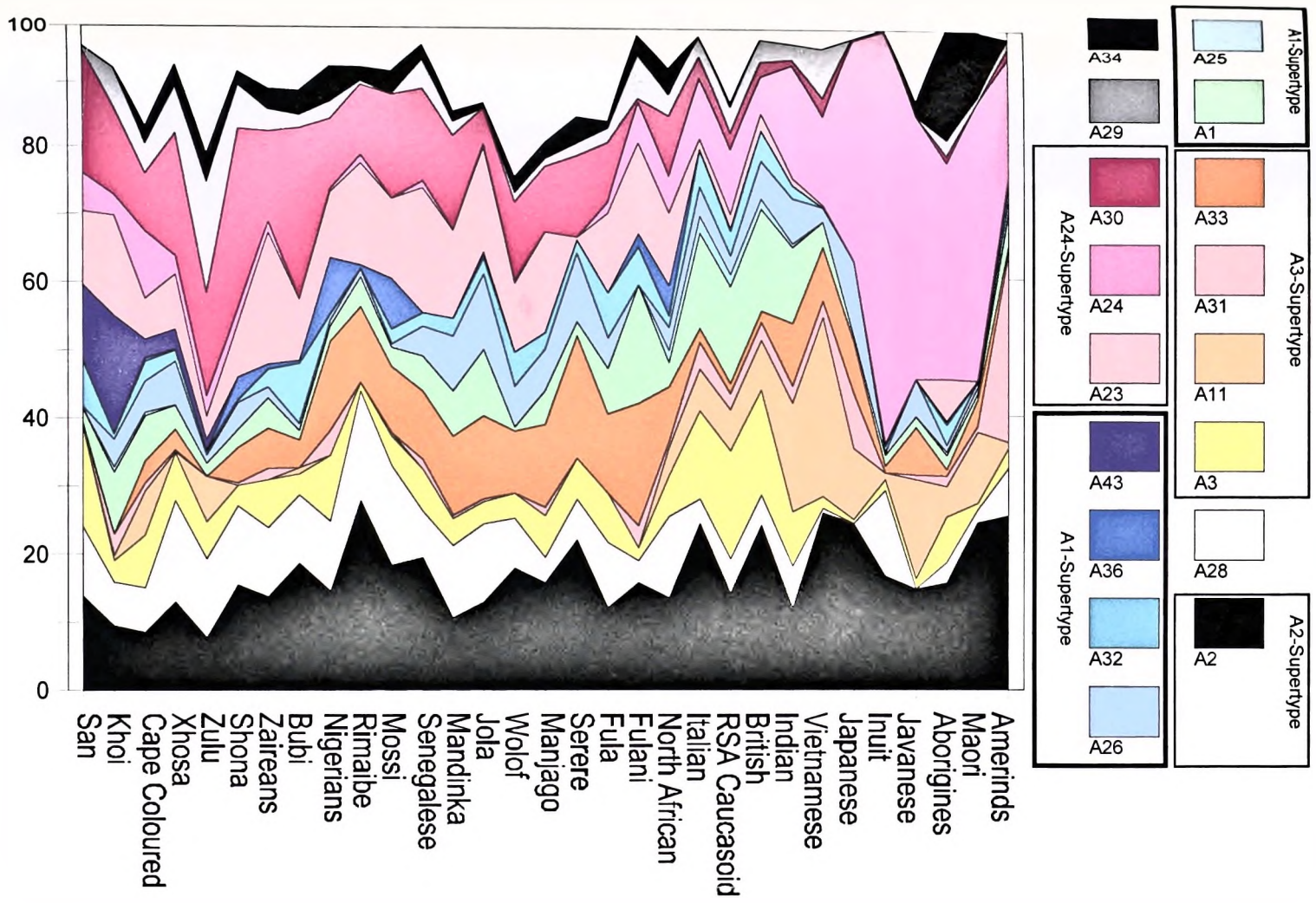


Figure 48. Percentage frequency of major HLA-A antigens (grouped and colour coded according to supertype) in selected populations (Table 16 gives details of references of publications used for population frequency data)

broad A28 allele is, however, better represented in the African populations and rare or absent from all other populations with the exception of the Inuits and to a lesser extent the Amerinds.

The two splits, A23 and A24, of A9 form, together with allele A30, the A24-supertype (red in Figure 48). This allele is mostly represented by A23 in African populations but by A24 in other populations. The southern African populations, as well as the Fula, Fulani and North Africans, also have moderate amounts of A24 but this split is extremely rare or absent in West Africa. A24 is, however, found very frequently in the Far East and the Pacific ranging from 31.8 percent in Australian aborigines to 63.8 percent in the Inuit population. A30 is well represented in all the African populations with the exception of the Fulani and is more frequent in southern and central Africa than in West Africa. Although grouping by supertype tends to average out the frequency of the particular phenotype being examined, the A24-supertype still shows a greater prevalence in the East and is more rare in Caucasoids with an intermediate prevalence in Africans.

The Caucasian populations have higher frequencies of A2 than most Africans but have considerably higher frequencies of A1 and A3 when compared with the African ethnic groups. Alleles A1 and A3 are the prototypes of two supertypes of the same name. The A33 specificity of the A9 allele is part of the A3 supertype and is most prominent in West Africa. South Africans exhibit higher A29 but this A9 specificity has not yet been assigned a supertype. The A31 specificity of A9 is very high only in the Amerinds and has also been classified as supertype A3. The San have a notably higher frequency of A3 than other ethnic groups originating in Africa. Only the Zulu and Cape Coloured in the African populations have A11, they share a 6 to 7 percent frequency of this allele with the Caucasians. The Asians have the highest frequency of the A11 allele ranging from 10.4 percent in the Japanese to 26.3 percent in the Vietnamese.

The broad A10 allele is made up of the splits A25, A26, A34 and A66. The two former alleles, A25 and A26, together with A32, A36, A43, A80 and A1 itself form the A1 supertype. A26 separates the Western Atlantic linguistic groups from the Bantu of West Africa. A25 is rare or absent from all the groups with a 1.4 to 2.4 percent incidence in the Caucasoids. The A43 allele which is most prevalent in the Khoisan was discovered in the

San (du Toit *et al.*, 1988) and is only present in southern Africa. The Australian aborigines and the Maoris share and are alone in having a high A34 (also part of the broad antigen A10) frequency.

Both the Fulani and the Cape coloured show relevant differences for most alleles compared to their geographically neighbouring populations. Apart from their lower A30 the Fulani also have a lower A28 and a higher A1 and A33. The Cape coloureds have a high A43 and A24.

8.3.2 Class I HLA-B comparison

Figure 49 shows that there are differences between southern, central and western African populations at the HLA-B locus. Again the Fulani have a unique frequency distribution and the Cape coloureds have many low frequency alleles. The Bantu populations have a higher incidence of B42, which has not yet been assigned a supertype, than the Khoisan and West Atlantic populatons and this allele is rare or absent from the non-African populations as well as the Fulani and the Serere. B8 which has also not been supertyped is common in Caucasoids and the West Atlantic groups and rare or absent in West Africa. Earlier sources of data group all the splits of B15 together into the broad B15 allele even though phenotypically they span four of the five supertypes of HLA-B, namely B7, B27, B58 and B62. Data is also missing for several of the HLA-B alleles making colour coding and grouping of the supertypes less meaningful I have, however, still attempted to present the data in this way since it does simplify the analysis (Figure 49). The B70 antigen, which is also a part of the broad B15 allele, is present in, and characteristic of, all the African populations.

The B58-supertype is made up of B*1516 and B*1517 (not shown see above) as well as the broad B17 allele which can be split into B57 and B58. This supertype is thus shown as a single band, the second series white band, in Figure 49. The San and the Khoi have very high B17 frequencies but the allele is rarely found in Asia.

The B62-supertype is also poorly represented on the graph as it is made up of numerous splits of B15 (including B*1501 the prototype B62 of the group), B46 (no data available), B13 and B52. Except for a high frequency of B13 in the Zulu and the Bubi, the B13 allele is

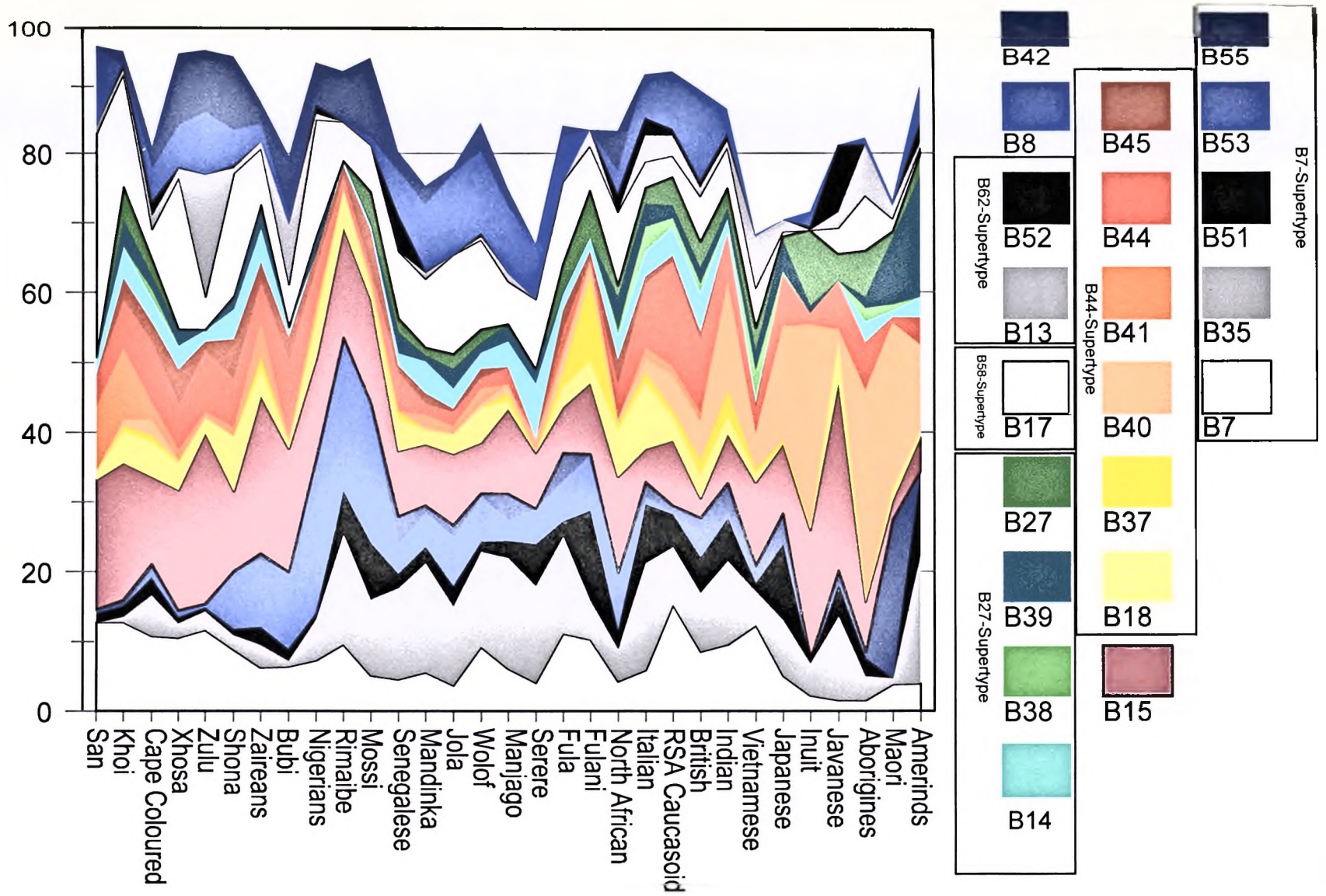


Figure 49. Percentage frequency of major HLA-B antigens (grouped and colour coded according to supertype) in selected populations
(Table 16 gives details of references of publications used for population frequency data)

rare globally with moderate frequencies in the Caucasians.

The B27-supertype is made up of B*1401 and B*1402, various splits of B15, B*2701-8, B38 and B39 as well as B48 and B73 for which data was incomplete. B27 is most prevalent in the Inuits while B39 is frequent in the Amerinds. B14 and B38, are rare or absent in the Bantu of West Africa, the Japanese, Inuits and Javanese and rare in the other populations.

The B44-supertype is better represented in Figure 49 with separate figures available for all of the supertype alleles except B*4901 and B*5001. The West Africans have less of this supertype than other Africans including the Fula and Fulani. The B12 allele has a contribution from both B44 and B45 in Africa but is mainly represented by B44 in other populations. The San and the Khoi have very high B41 frequencies. The Fulani have a very high B37 frequency. B40 has a high frequency in the East and Oceania with the exception of the Inuits and is rare in the Caucasians.

Of the many alleles included in the B7-supertype [Table 17 (c)] separate data is only available for B7, B35, B51, B53 and B55. B35 and B51 are rare in southern Africa but well represented in all other populations with the exception of the Maori. B53 is notably absent from southern Africa and Asia and rare in the Caucasians but common in the other Bantu and West Atlantic linguistic groups of central and West Africa. B55 is also absent or rare in southern and central Africa and best represented in the Maori. The prototype allele, B7, is however present in all the populations shown including the southern Africans.

The distribution of the supertypes of HLA-B will be clearer when more high resolution data becomes available.

8.3.3 Class II HLA-DR comparison

Although there is also an overlap in the peptide binding repertoires of class II HLA molecules, the fit in these molecules is less rigid making it difficult to define distinct supermotifs and supertypes. However, the four African linguistic groups (represented by the San, the 7 Bantu tribes, the Senegalese and the Algerians) have distinct characteristics at this locus (Figure 50). In Africa DRB1*02 (particularly represented by DRB1*15) is found most frequently in central Africa (with maxima of 20.7 and 18.4 percent in Zaireans and the

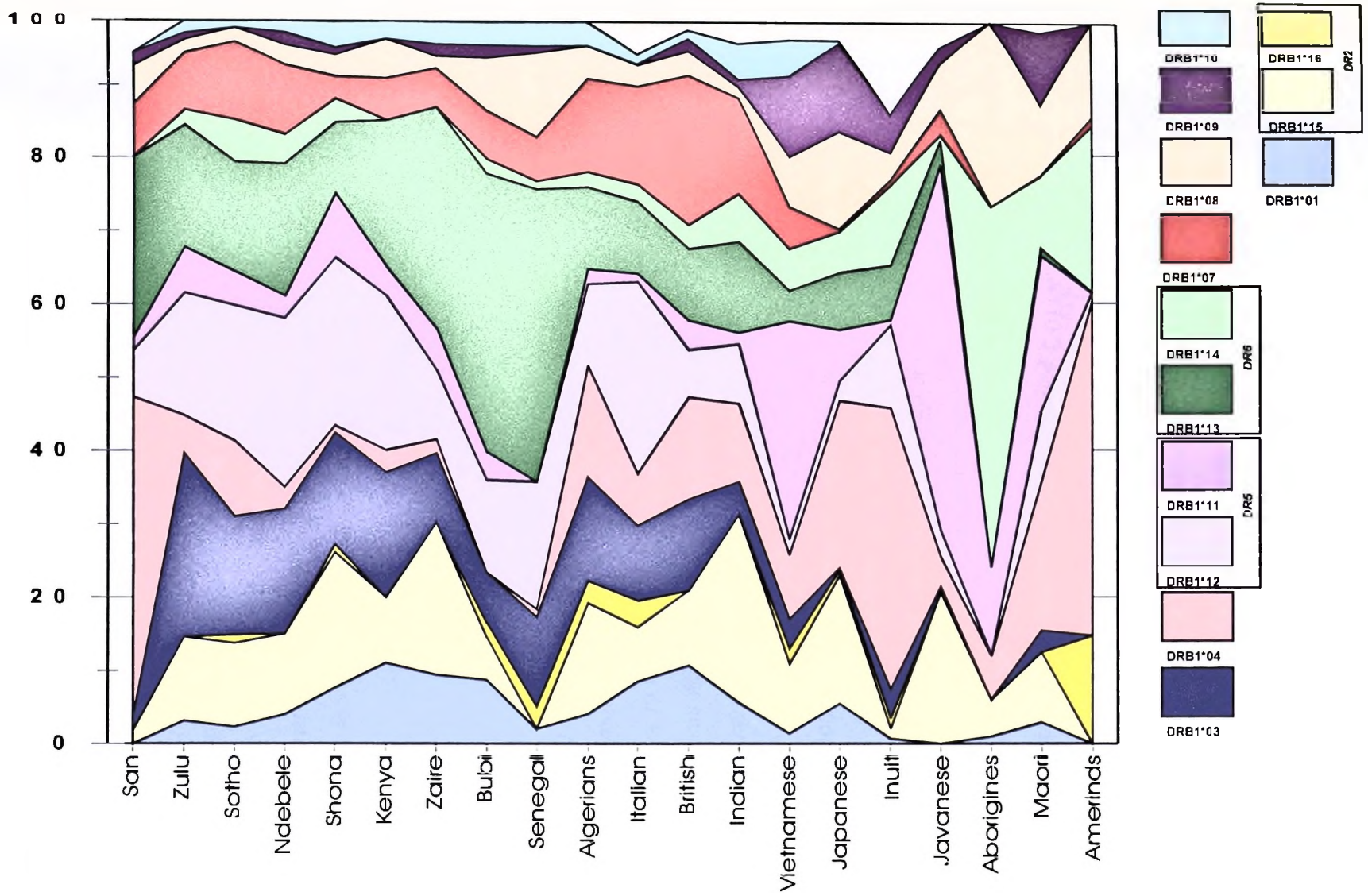


Figure 50. Percentage frequency of major HLA-DR antigens in selected populations (Table 16 gives details of references of publications used for population frequency data)

Sinona, respectively). The southern Africans all had about 11 percent of this allele which is rare in the Bubi and Senegalese. The Senegalese have only DRB1*16 and this allele is also present exclusively and extensively in the Amerinds. DRB1*03 is frequent in all African and Caucasian populations with the exception of the San but rare or absent in the East and Oceania. DRB1*10 has a low frequency in most populations and is completely absent in the San and in Oceania. Its highest frequency (5.6 to 5.7 percent) is in Asia. DR5 is present in all populations with the exception of the Amerinds but is represented by DRB1*11 in Africans, Caucasians and the Inuits but by DRB1*12 in Asia and Oceania. DRB1*07 is rare only in Asia and Oceania and present most frequently in Caucasoids. DRB1*04 is rare in all the Africans except the Sotho (10.3 percent) and the Algerians (15.1 percent) but is present in 43.5 percent of the San. It is present in 4 to 14 percent of other populations but found frequently in the Japanese (22.8 percent), Inuits (38.3 percent) and the Amerinds (45.4 percent). The DRB1*13 split of DR6 is particularly frequent in Central and West Africa and to a lesser extent in southern Africa, north Africa, Europe and Asia but rare or absent in Oceania. Conversely the DRB1*14 split is rare or absent in most populations with the exception of the Aborigines (49.0 percent) and the Amerinds (22.7 percent). DRB1*08 is represented in all the populations but has a greater, though still intermediate frequency, in Senegal and Japan but a high frequency among the Australian aborigines. DRB1*09 is rare except in Asia and among the Maoris. The West Atlantic population of Senegal varies from the rest of Africa at many of the DR loci. The HLA 1991 frequency tables list DRB1*01 as absent in the San while du Toit *et al.*, 1990 comment on its frequency (14.9 percent); the latter is more likely to be correct as du Toit is responsible for most frequency figures for this population.

8.3.4 Class II HLA-DQ comparison

The DQ5 split of DQB1*01 is evenly distributed in most populations but absent in the Amerinds (Figure 51). Its highest frequency (46.5 percent) is in the Australian Aborigines. The DQ6 split of this allele is also well represented in most populations with the exception of the Amerinds but is also less common in the Senegalese and the Vietnamese. DQ2 is frequent in African and Caucasoid populations but rare in Asia and Oceania. The DQ7 split of DQB1*03 is well represented in all populations except the Italians and the Japanese. DQ8 is seen less frequently except in the Maori and the Amerinds and to a lesser extent the San. It is absent in the Bubi and the Australian Aborigines. Finally, the DQ9 split is seen most

frequently in the Vietnamese but is absent in the Italian, Indian, Aborigines and Amerinds. DQ4 is seen in the southern and central African populations and in Oceania. It is most frequent in Japan but rare in West and North Africa and in Caucasoids.

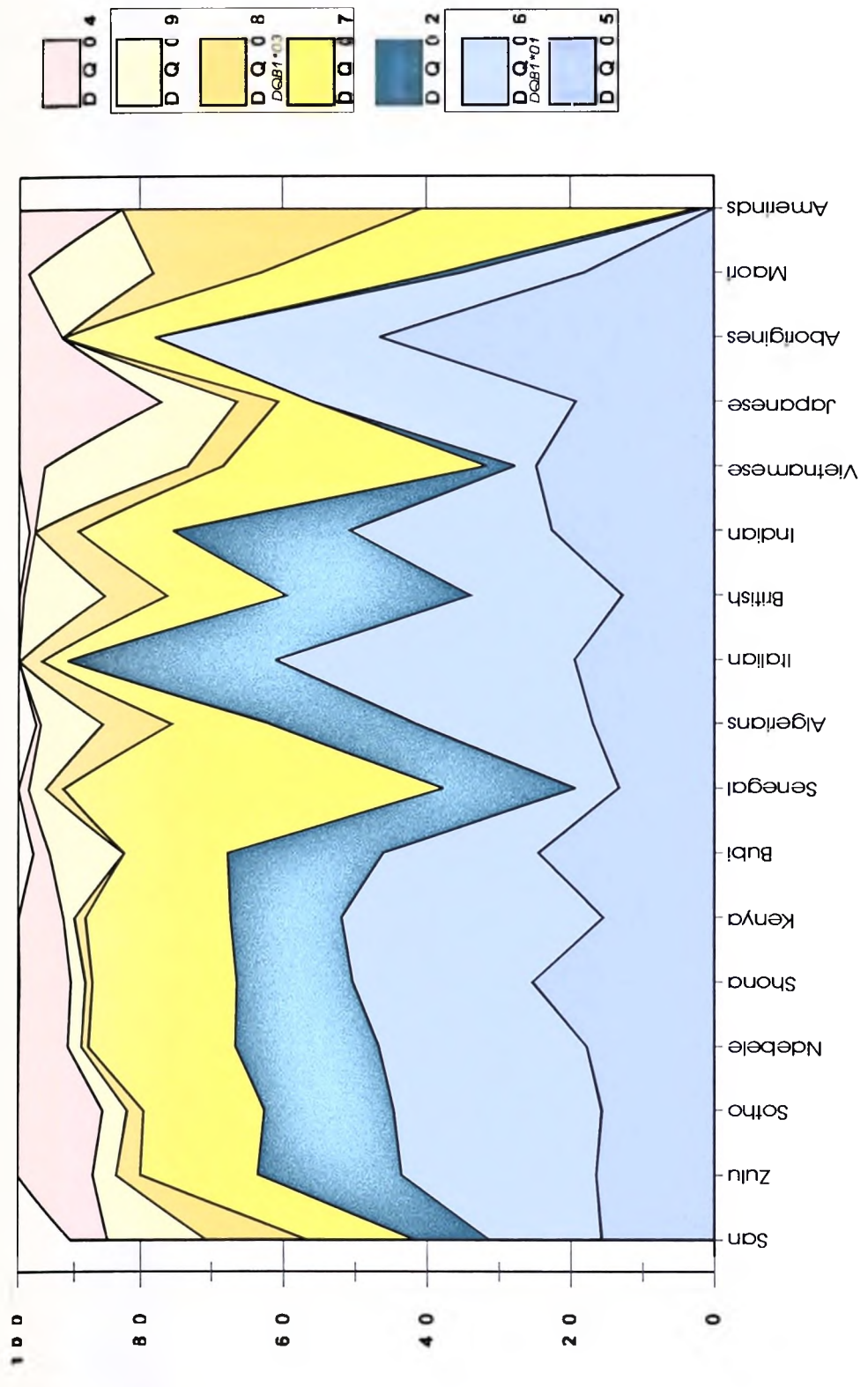


Figure 51. Percentage frequency of major HLA-DQ antigens in selected populations (Table 16 gives details of references of publications used for population frequency data)

8.3.5 Local HLA-DR typing results

Table 18 shows the results of the HLA-DR typing of 13 local southern African specimens. Both low and high resolution SSP-PCR typing was performed for the more common specificities which both refined and confirmed initial results. All four of the acute patients had the broad, serologically defined, DR3 specificity which was also present in three of the chronic carriers and the low responder vaccinated individual. The second allele of the low responder was found to be DR7. One of the individuals acutely infected had two DR3 alleles, DRB1*0310 and DRB1*0316. The broad DR5 specificity, and the DR11 split in particular, was most common in the chronic carriers, being present in three of the 7 who were typable. Half (2/4) of the acute patients shared a DR4 allele while the fourth had the rare DRB1*1605 split of the DR2 allele. The healthy individual was the only specimen with the DR13 split of the DR6 specificity. The DR52 haplotype was by far the most common haplotype among the southern Africans typed representing at least one allele in 12 of the 13 specimens. The second allele was not discerned in 6 of the specimens.

Table 18. HLA-DR types determined for southern African specimens

	HLA-DR	Diagnosis
CH	DRB1*1307, 1311, 1314, 1321 or 1322 (DR52)	Healthy individual with no markers of HBV infection
SB	DRB1*0304-0306 (DR3) DRB1*0701 (DR7)	HBV low responder vaccinated individual
ES06	DRB1*0302 = DR18 DRB3 (DR52) DRB1*04 (DR53)	Acute HBV
ES15	DRB1*0301 = DR17 (DR52) DRB1*1605 = DR2 (DR51)	Acute HBV
ES16	DRB1*0310 = DR3 DRB3 (DR52) DRB1*0316	Acute HBV
HB05	DRB1*0311 = DR3 DRB3 (DR52) DRB1*0415 DRB4 (DR53)	Acute HBV
N231	DRB1*0302 = DR18 DRB1*1201 1206	Chronic carrier
N232	DRB1*0308	Chronic carrier
N234	DRB1*1120/21 (DR11 DR5)	Chronic carrier
N235	DRB1*1001	Chronic carrier
N236	DRB1*1121 *1131 *1134 *1136 (DR11 = DR5) DRB1*0801-0819 *0821 (DR8)	Chronic carrier
N237	DRB1*11011/2/3 *1103/4 *1106 *11081/2 (DR11 = DR5)	Chronic carrier
N238	DRB1*0304 (DR3)	Chronic carrier

8.4 Discussion

Phylogenetic analysis of ethnic groups has defined five world population groups: Africans, Oceanians (includes Australian and Taiwanese aborigines, Yakuts and Indians), North Asians (Northern China, Mongolians, Japanese, Koreans and Tibetans), Caucasoids and South-east Asians (Imanishi *et al.*, 1992). Twin studies and segregation analysis have shown that the clinical outcome of HBV infection is both variable and determined, at least in part, by the host's genetic makeup (Thursz, 1997a). It has been shown (Ferrari *et al.*, 1990) that T helper cell responses in patients with chronic HBV infection are attenuated by comparison to those in patients with acute, self-limiting infection and indeed numerous DRB1 and DQB1 loci have been found to be associated with the outcome of HBV infection, the vaccine response or the response to therapy (Table 19). To better understand the African host we have therefore compared the southern African population frequencies of the HLA antigens with population frequencies selected from the other world population groups. We also typed a small sample of local specimens (both infected and uninfected) for comparison with population statistics.

Antigens which are considered relatively "African specific" are A30, A23, B15 (represented by B70), B42, DR6 and to a lesser extent A28 as well as a low frequency or absence of A11 and A24 (Dard *et al.*, 1991). Southern Africans are defined by A43 and West Africans by A33. A43 was first identified in the San people and is unique to southern Africa (Nurse *et al.*, 1975; du Toit *et al.*, 1990). Frequency levels of the DRB1*13 split of DR6 are very high in Central and West Africa reaching a maximum of 39.8 percent in the Senegalese. The Senegalese also have 12.2 percent of DR8 which is rare or absent from other African and Caucasian populations and have none of the DR15 split of DR2 which is common in most other populations. DR3 and DR7 are rare in Asia and the Pacific but quite common elsewhere particularly in southern and central Africa. Although their numbers are small in southern Africa today, the San and the Khoi are a unique people. Until recently most alleles were considered effectively neutral but evidence is accumulating which indicates that the HLA alleles are in fact subject to considerable selection pressures (Allsopp et al., 1992). It has not yet been established whether the Bantu and the Khoisan are from common stock but if population haplotypes are due to selection, the Khoisan and the Bantu have been separated for a very long time. The San have no B42 and less DR2, DR3 and DR5 than the rest of the

Table 19. Published host susceptibility genes which modify the disease phenotype, the vaccine response or the response to therapy

	Exposure to HBV	Outcome of infection		Outcome of chronic infection		Outcome of interferon therapy	Outcome of vaccination		
	Infection	Acute infection	Chronic infection	asymptomatic	CAH	HBV DNA clearance	Good response	Low response	Non-response
HLA-A1									+ (10)
HLA-A2								+ (100)	
HLA-A19							+ (10)		
HLA-B5							+ (10)		
HLA-B7								+ (10)	
HLA-B8									+ (10)
HLA-B*5				+ (4)			+ (10)		
HLA-B27							+ (10)		+ (10)
HLA-B35					+ (8)				+ (7)
DPB1*04							+ (3)		
DPB1*1101									+ (3,)
HLA-DQ1					- (13)		+ (3; DQB1*0501); + (5; DQB1*0602/3)		- (5; DQB1*0604)
HLA-DQ2									+ (3, 10)
HLA-DQ3					+ (11, DQB1*0301)		+ (3; DQB1*0301)	+ (6)	
HLA-DR1							+ (3; DRB1*010)		
HLA-DR2	- (1)	- (1)	- (1)				+ (7; DR15)		
HLA-DR3					+ (9)	+ (9)			+ (2; 3; 10)
HLA-DR4								+ (10)	
HLA-DR5							+ (3)		
HLA-DR6		+ (5; 12; DRB1*1301/2)	- (5; 12; DRB1*1301/2)			+ (9)	+ (7; DRB1*1301)	+ (5, DR14)	+ (7; DRB1*1302)
HLA-DR7			+ (1)						+ (2; 3; 14)

Numbers in brackets refer to reference footnote overleaf; - indicates that the allele is protective and is therefore under-represented in patients or vaccinees; + indicates that the allele enhances susceptibility to a particular disease (or therapy) outcome and is therefore found more frequently in patients with this disease (or therapy) outcome or response; CAH = chronic active hepatitis

southern African populations (Table 20). They also have more A43, B8, B41, Cw6, DR1 and considerably more DR4 (du Toit *et al.*, 1990). They are thought to have contributed the southern African component of the coloured peoples of the Western Cape (Hammond *et al.*, 1997) and also have all three of the so-called "European" marker antigens, A3, B7 and B8. The Xhosa have had the greatest contact with the San (Jenkins *et al.*, 1970). The Bantu and the West Africans are genetically closely related in keeping with the fact that they both belong to the Niger-Congo Phylum. The non-Bantu West Africans of Senegal, Mali and The Gambia are separate from the Bantu and the San and the Fulani are most closely related to Caucasians.

The similarity of the HLA antigen frequencies of the four Bantu populations, the Sotho, Zulu, Ndebele and Shona and their divergence from the San can be seen in Table 20. The four most frequent alleles include two protective alleles, the DRB1*15 split of DR2, and the DRB1*13 split of DR6. The other two common alleles are DRB1*03 and the DRB1*11 split of DR5.

A number of genetic and environmental factors determine disease outcome after exposure to HBV (Thursz, 1997a). Several class II antigens have been found to be significantly less

Table 20. Population frequencies of the HLA-DR alleles in four Bantu populations and the San people (Table 16 gives details of references used for population frequency data)

DRB1	San	DRB1	Sotho	DRB1	Zulu	DRB1	Ndebele	DRB1	Shona
DRB1*01	0	DRB1*09	0	DRB1*16	0	DRB1*16	0	DRB1*09	1.08
DRB1*16	0	DRB1*10	1.149	DRB1*09	1.04	DRB1*10	2	DRB1*04	1.08
DRB1*14	0	DRB1*16	1.149	DRB1*10	2.08	DRB1*09	2	DRB1*16	1.08
DRB1*10	0	DRB1*01	2.29	DRB1*08	2.08	DRB1*12	3	DRB1*08	3.26
DRB1*15	1.9	DRB1*08	2.29	DRB1*14	2.08	DRB1*04	3	DRB1*14	3.26
DRB1*03	1.9	DRB1*12	4.59	DRB1*01	3.125	DRB1*08	3	DRB1*07	3.26
DRB1*12	1.9	DRB1*14	5.74	DRB1*04	5.208	DRB1*14	4	DRB1*10	4.34
DRB1*09	1.9	DRB1*04	10.34	DRB1*12	6.25	DRB1*01	4	DRB1*01	7.6
DRB1*08	5.6	DRB1*07	11.49	DRB1*07	8.33	DRB1*07	10	DRB1*12	8.69
DRB1*11	6.3	DRB1*15	11.49	DRB1*15	11.458	DRB1*15	11	DRB1*13	9.78
DRB1*07	7.3	DRB1*13	14.94	DRB1*13	16.66	DRB1*03	17	DRB1*03	15.21
DRB1*13	24.5	DRB1*03	16.09	DRB1*11	16.66	DRB1*13	18	DRB1*15	16.47
DRB1*04	43.5	DRB1*11	18.39	DRB1*03	25	DRB1*11	23	DRB1*11	22.8

1 Almarri *et al.*, 1994
2 Craven *et al.*, 1986
3 Desombere *et al.*, 1998
4 Giani *et al.*, 1979
5 Hohler *et al.*, 1997
6 Hsu *et al.*, 1996
7 Lango-Warensjo *et al.*, 1998
8 Mota *et al.*, 1987
9 Scully *et al.*, 1990
10 Stachowski *et al.*, 1995
11 Thio *et al.*, 1999
12 Thursz *et al.*, 1995
13 van Hattum *et al.*, 1987
14 Walker *et al.*, 1981

common in chronically infected patients than in either healthy controls or in healthy immune patients (Table 19). A study conducted in the Gambia (Thursz *et al.*, 1995), which was subsequently confirmed in a European cohort (Hohler *et al.*, 1997), associated DRB1*1302 with the clearance of HBV infection. This allele is among the four most common HLA types in all four Bantu populations listed in Table 20, having a frequency of 14.9 and 16.7 percent in the Sotho and Zulu populations, respectively. Supporting its protective role, none of our HBV patients were found to have this allele. Another study in Qatar associated DR2 with disease clearance and DR7 with persistent infection (Almarri *et al.*, 1994). The DR15 split of DR2 is also common in the Bantu populations, present in 11.4 percent of both the Sotho and the Zulu, but this protective allele was not found in any of the patients tested.

Persistent infection is associated with histologically defined liver damage and raised liver transaminases (chronic active hepatitis, CAH). The duration of the active phase of persistent infection is variable and is superceded by asymptomatic HBsAg carriage when viral replication is under control, transaminases normalize and liver injury improves. Dusheiko *et al.* (1985) found that more than 90 percent of adult South African carriers had less than a twofold elevation of serum aminotransferases indicating that they had progressed to the predominantly non-replicative phase of the disease. Control of viral replication is mediated by CD4+ and CD8+ T-cells responding to HBV antigens presented by HLA class II and class I molecules, respectively, and therefore it is not surprising that various HLA alleles have been associated with the severity of disease in chronic carriers. Despite the fact that CAH is not a common disease manifestation in this country, DR7 is present in 11.5 percent of the Sotho and 8.3 percent of the Zulu and DQ7, also associated with CAH, is present in approximately 17 percent of both the Sotho and the Zulu. We found DR7 in the specimen from the vaccinated individual only. HLA–B35 which has also been associated with CAH is however a rare allele in southern and central Africa while one of the definitive African alleles, B15, has been associated with asymptomatic HBsAg carriage. B15 is present in 17.3 percent of Xhosa and 24.5 percent of the Zulu. The DQ1 allele which is found less frequently in CAH is also the dominant allele in all the Bantu populations with a frequency of 40 to 50 percent. Different HLA disease associations in different population groups is however to be expected because of variations between population allele frequencies (Thursz, 1997a).

Ten to 14 percent of Caucasian populations, and even more in the Far East, fail to mount an

adequate antibody response after immunization with HBsAg vaccines. Many studies have shown a link with HLA alleles or haplotypes (Table 19). For example DQB1*0602 and DQB1*0603 were found to be associated with vaccine responses while DQB1*0604 was associated with non-responsiveness. Similarly, DRB1*1301 and DR15 were amongst the responder haplotypes while DRB1*1302 was a non-responder haplotype. The low responder specimen from the vaccinee which we typed had both DR3 and DR7, both of which have been associated with non-responsiveness. Antibody production is normally supported by an MHC class II restricted T-helper cell response and therefore it is not surprising that failure to respond was frequently associated with MHC phenotype. The interaction between the HLA molecules is also important, however since there were several HLA-identical responders and non-responders, other, and not necessarily genetic, factors must also predispose the HBsAg response (Lango-Warensjo *et al.*, 1998).

Although CD4+ T-cell responses (Jung *et al.*, 1995) and CD8+ CTL responses (Giani *et al.*, 1979) increase substantially during seroconversion to antibody to HBeAg, Scully and colleagues (1990) found only DR3 and DR6 influenced host response to interferon. The common African allele, DRB1*1302, which is associated with acute hepatitis and disease clearance belongs to the DR6 'supertype' and thus may also be important for viral clearance in response to interferon treatment.

Most of the patients had the common non-protective haplotypes DR3 (ES06, ES15, ES16, HB05, SB, N231, N232 and N238) and DR11 (N234, N231, N236 and N237). DR3 has associations with more active disease and was found in all four acute patients as well as three of the chronic carriers. Two of the latter appeared to be homozygous for DR. Heterozygosity at the DR and DQ Class II loci was also found to be associated with protection from persistent infection (Thursz, 1997b), while homozygosity for HLA class I (Pollicino et al., 1996) and class II (Thio *et al.*, 1999) was found to be responsible for an insufficient response to HBV, favouring chronic outcome of the infection and long-lasting persistence. Only one allele (which could indicate homozygosity) was detected in 4/7 chronic carriers tested.

CHAPTER 9

MAPPING THE REGIONS OF THE HBV CORE GENE PRODUCT WHICH ELICIT AN IMMUNE RESPONSE IN ACUTE AND CHRONIC PATIENTS FROM SOUTHERN AFRICA

9.1 Introduction

The cellular immune response to pathogens depends on the presentation and recognition of their protein antigens in the form of intracellularly processed peptides bound to major histocompatibility complex (MHC) molecules expressed on the cell surface. Peptides, 8 to 10 amino acids in length, derived from antigens synthesized in the cytoplasm, are presented in conjunction with MHC class I molecules while 10 to 20 mer peptides derived from exogenous antigens are usually presented in the context of MHC class II molecules. Several methods have been employed to locate these T-cell epitopes within the amino acid sequences of viral and bacterial protein antigens. One common approach is to synthesize overlapping peptides spanning the entire sequence of a protein antigen and then test them for their capacity to stimulate T-cell proliferative or cytotoxic responses *in vitro*. These methods are both cost– and labour-intensive and computer-based approaches are becoming more popular. Used prior to proliferative or CTL assays their predictions can assist in a more cost effective choice of peptides for empirical studies and aid in interpretation of the assay results.

Although there is a lack of helicity in peptides bound in the MHC groove, they have been found to bind as a ß-strand with a 130° twist in which the hydrophobic anchor residues have a spacing and periodicity similar to the amphipathic helix (Stern *et al.*, 1994). Thus one algorithm for epitope prediction searches a protein's primary structure for peptides with a high probability of folding as amphipathic structures. The AMPHI algorithm (Margalit *et al.*, 1987) is based on this principal and was originally found to have a predictive power of 70 percent. Since a large proportion of peptides that bind to a specific MHC allele have specific characteristics or favoured HLA-binding motifs, another approach is to search the primary protein sequence for regions that contain specific anchor residues. Programmes which use these algorithms alone or in combination have been demonstrated to accurately

and efficiently predict T-cell epitopes (Meister *et al.*, 1995).

Two such prediction algorithms available on the internet are SYFPEITHI (Rammense *et al.*, 1999; http://134.2.96.221) and BIMAS (BioInformatics and Molecular Analysis Section; Parker *et al.*, 1994; http://bimas.dcrt.nih.gov). SYFPEITHI searches the primary protein structure for peptides which match those within a database of published motifs taking into consideration all frequent amino acids, particularly those in anchor and auxiliary-anchor positions of the motif, and then assigns a value to the peptide which reflects its binding ability with a given allele. BIMAS, developed at the NIH (National Institutes of Health, Bethseda, Maryland, USA), ranks potential 8-, 9- or 10-mer peptides based on their predicted half-time of dissociation with HLA class I molecules. It is based on coefficient tables deduced from the published literature by Dr. Kenneth Parker. The scores assigned to different peptides by these algorithms differ considerably but, in general the same peptides are ranked within the top 10 or 20 by both programmes and can be used to calculate the regions of a protein which have the highest binding motif density.

I used computer prediction and the T-cell proliferation assay to determine immunogenic regions of the HBV core gene product. This region was chosen because changes in the immune recognition sites of the core antigen have been found to illicit (or evade) immune clearance and affect the outcome of infection (Ferrari *et al.*, 1988).

9.2 Materials and Methods

9.2.1 Study Populations

Four individuals (three males and one female) who were diagnosed with acute hepatitis B in Johannesburg hospitals and 8 chronic carriers from the North East province of Namibia, all of whom had seroconverted to anti-HBe, were selected for *in vitro* epitope studies. The diagnosis of acute hepatitis was based on the detection of IgM anti-HBc antibodies in the serum, the recent onset of jaundice and other typical symptoms of acute hepatitis.

9.2.2 Serology

Commercially available radioimmunoassay test kits were used in accordance with the manufacturer's instructions for the determination of the HBsAg, anti-HBs, HBeAg, anti-HBe, anti-HBc IgM and anti-HBc IgG (Abbott laboratories, North Chicago, IL).

9.2.3 Genotype determination

The HBV genotype of the 12 patients was determined using the RFLP screening assay or HBV genotype EIA, or both, as described in Chapter 4.

9.2.4 Low resolution HLA-DR typing of southern African patients

The HLA-DR type was determined for all patients using low– and/or high-resolution sequence specific PCR (SSP) HLA typing kits from Dynal (Norway) as described in paragraph 3.3.5 and Chapter 8.

9.2.5 Lymphocyte isolation and the proliferation assay

Cell preparation and culture conditions have been previously described (paragraph 3.3). Peripheral blood mononuclear cells (PBMCs) were obained from EDTA venous whole blood by Ficoll-Paque (Pharmacia, Piscataway, NJ) gradient separation as described in paragraph 3.3.1. The PBMCs were then exposed to different concentrations (0.01 and 2 μg/ml) of 17 overlapping peptides covering the entire core protein as described in paragraph 3.3.2, and the level of stimulation elicited by different regions of the core protein measured by pulsing the cells with ^{3}H-thymidine on day five and measuring the incorporation of the ^{3}H-thymidine.

9.2.6 Peptide preparation

Branched or monomeric peptides were prepared at the Glasgow Institute of Immunology. They were a kind gift from Dr. William Carman. Their sequence, re-hydration and dilution is described in paragraph 3.3.2.

9.2.7 Expression of data

The arithmetic mean of counts per minute (cpm) from quadruplicate cultures containing a given dose of peptide were recorded. The value of the peptide dose yielding the greatest response was recorded as the peak response. The peak response data for a given antigen is expressed as the arithmetic mean ± standard deviation (SD), as well as the aritmetic mean ± the standard error (SE). The 95 percent confidence interval was also listed (Appendix E, Table 27b). The stimulation index (SI) is the ratio of the cpm in the presence and absence of antigen. An SI of greater than 2.5 was regarded as a positive response.

9.2.8 Epitope prediction

The SYFPEITHI programme was run from the internet and the output copied to file to provide a listing of the top ten predicted epitopes for the MHC class I and class II alleles available at this site (HLA-A1, A*0201-0203, A26, and A3; HLA-B*0702, B8, B*1510, B*2705, B*2709 and B*5101 and HLA-DRB1*0101, DRB1*0301, DRB1*0401) and their predicted binding scores. Similarly the BIMAS program (Rammense *et al.*, 1999) was run and the top 10 binding motif sites for each of the following alleles stored: HLA-A*0205, A24, A*1101, A*3101, A*3302 and A68.1.; and HLA-B*14, B*2702, B*3501, B*3701, B*3801, B*3901-3902, B*40, B*4403, B*5102-5103, B*5201, B*5801 and B*60-62. A histogram of motif density at each core amino acid was then drawn.

9.3 Results

9.3.1 Genotypes

Two of the four HBV isolates from the patients with acute HBV were found to be from genotype D and the remaining two were from genotype A. Seven of the eight chronic carriers were infected with HBV from genotype E while one, N236, was infected with HBV from genotype D (Table 21).

Table 21. Genotype and HLA-DR type of specimens chosen for proliferation analysis

	HLA-DR	Genotype	Diagnosis
ES06	DRB1*0302 = DR18 DRB3 (DR52); DRB1*04 (DR53)	A	Acute HBV
ES15	DRB1*0301 = DR17 (DR52) DRB1*1605 = DR2 (DR51)	D	Acute HBV
ES16	DRB1*0310 = DR3 DRB3 (DR52); DRB1*0316	D	Acute HBV
HB05	DRB1*0311 = DR3 DRB3 (DR52) DRB1*0415 DRB4 (DR53)	A	Acute HBV
N231	DRB1*0302 = DR18; DRB1*1201-1206	E	Chronic carrier
N232	DRB1*0308	E	Chronic carrier
N233	NOT DETERMINED	E	Chronic carrier
N234	DRB1*1120/21 (DR11 - DR5)	E	Chronic carrier
N235	DRB1*1001	E	Chronic carrier
N236	DRB1*1121 *1131 *1134 *1136 (DR11 = DR5); DRB1*0801 0819 *0821 (DR8)	D	Chronic carrier
N237	DRB1*11011/2/3 *1103/4 *1106 *11081/2 (DR11 = DR5)	E	Chronic carrier
N238	DRB1*0304 (DR3)	E	Chronic carrier

9.3.2 HLA-DR typing

All four of the acute patients had the broad, serologically defined, DR3 specificity which was also present in three of the chronic carriers. One of the acute patients, patient ES16, had two DR3 alleles, DRB1*0310 and DRB1*0316. One of the chronic carriers, N233, could not be DR-typed since genomic DNA from this patient did not amplify in the SSP PCR. It is possible that the DNA pellet was lost during chromosomal DNA extraction. The DR11 split of DR5 was most common in the chronic carriers being present in three of the 7. Two of the four acutes, ES06 and HB05, also shared a DR4 allele. The fourth acute patient, ES15, had the rare DRB1*1605 specificity of the DR2 allele. The DR52 haplotype represented at least one allele in 10 of the 11 patients typed in this study. The DR types are listed in Table 21.

9.3.3 Regions predicted to contain the HBV linear epitopes

The motif density at each core amino acid position, obtained by mapping the position of the top 10 predicted class I and class II core epitopes for the 38 unique alleles available at the time of computation, produced the histogram shown in Figure 52. The 7 separate regions of most dense motif clustering included the genotype G 12 amino acid insert which starts at

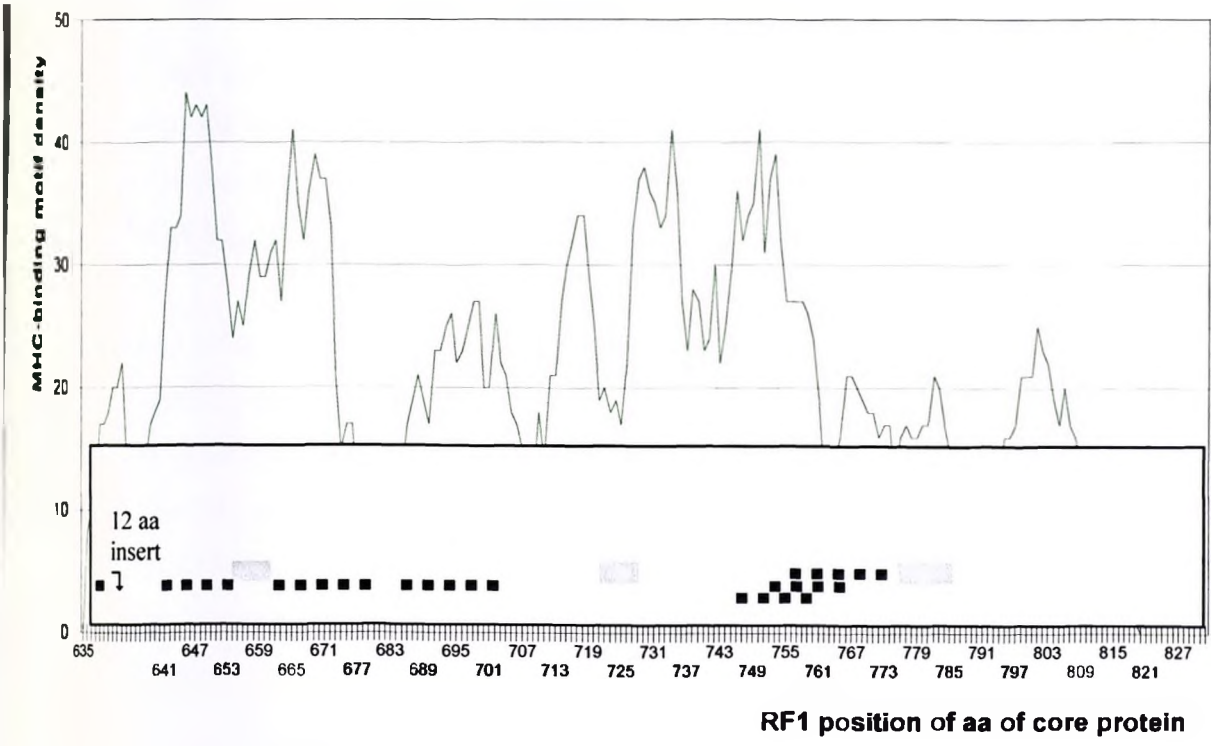


Figure 52. Published epitopes in the HBV core protein superimposed on the histogram of MHC-binding motif density as predicted by BIMAS and SYFPEITHI. Published Th epitopes ▪▪▪▪ and CTL epitopes are marked as shown

core residue three as well as the following 6 regions (amino acids are numbered from the start of RF1) which will be referred to as the first to sixth motif-rich region: 638-672, 684-705, 709-761, 765-774, 778-785 and 798-802, respectively. Since binding motifs tend to cluster together (Meister *et al.*, 1995), these 7 regions are predicted to contain the T-cell linear epitopes of HBV. Table 22 shows the specific alleles predicted to bind in each of these regions.

9.3.3.1 The motif-rich regions in relation to published class I epitopes

The three published class I epitopes extend from RF1 amino acid 652-661 (core-18-27) within the first motif-rich region, RF1 amino acid 721-734 (core-87-100) in the third motif-rich region and RF1 amino acid 775-785 (core-141-151) which coincides with the fifth motif-rich region. These epitopes are reportedly HLA-A2, A11 and A31/68 specific, respectively. The only HLA-A2 binding motif present within the first published epitope is specific for the HLA-A0201 allele. Eight of the top 10 predicted epitopes for the allele HLA-A*1101 were found within the region RF1 amino acid 719-746 although predicted binding values at these sites were low. Predicted binding motifs for both A*3101 and A68.1 were found within the region in which the third class I published epitope was identified.

9.3.3.2 The motif-rich regions in relation to published class II epitopes

Two of the 6 published class II epitopes occur in the first motif-rich region. The first of these covers the first 20 amino acids of the core gene, RF1 amino acid 635-654 (core-1-20). The second published epitope, RF1 amino acid 662-681 (core-28-47), is specific for DR7 and DP0402 and extends beyond the first motif-rich region.

The second HLA motif-rich region extends from RF1 amino acid 684-705 and coincides almost exactly with the published epitope core-50-69 (RF1 amino acid 684-703). The remaining three class II published epitopes overlap one another. The first is DR1 specific and extends from RF1 745-759 (core-111-121). The last two class II epitopes extend beyond the third motif-rich region. The first of these is DR52/DR6 specific and extends from RF1 amino acid 751-765 (core-117-131) and the last, also DR1 specific, extends from RF1 amino acid 754-773.

9.3.4 Proliferation responses to the core peptide analogues

The class II DR *in vitro* responses to the 17 overlapping core peptide analogues in PBMCs

Table 22. HLA-binding motifs present in predicted and published epitopes

Region of densest motif		G insert	638-672			684-705	709-761			765-774	778-785	798-802
Position of Published			635-654	652-661	662-681	684-703	721-730	745-759	751-765	754-773	774-785	
Published HLA			ND	A2	DR7/	ND	A11	DR1	DRw52/DR6	DR1	A31/	
Motifs present	A1	●	●	●	●	●		●	●	●	●	
	A*0201	●	●	●	●	●	●	●	●	●	●	
	A*0202		●		●	●				●		
	A*0203		●		●	●	●		●	●		
	A*0205	●	●					●	●	●		
	A3	●	●		●	●	●	●			●	
	A*1101						●		●	●	●	
	A24		●	●	●		●	●	●			
	A26	●	●	●	●			●		●	●	
	A*3101						●		●	●	●	
	A*3302					●	●		●	●	●	●
	A68.1						●	●	●	●	●	●
	B*0702		●	●	●	●	●			●		
	B8	●			●	●	●					
	B14	●	●		●	●	●	●		●		
	B*1510		●		●					●		
	B*2702	●	●				●	●		●		●
	B*2705	●					●	●		●		●
	B*2709	●	●				●			●		●
	B*3101											
	B*3501		●	●	●							
	B*3701		●		●			●				
	B*3801		●		●	●						
	B*3901	●	●		●	●	●	●		●		
	B*3902		●		●	●		●		●		
	B40		●		●			●				
	B*4403		●				●	●				
	B*5101	●	●	●	●	●						
	B*5102		●	●		●		●	●	●		
	B*5103		●	●	●	●		●				
	B*5201		●	●				●	●		●	
	B*5801	●	●			●			●	●	●	
	B60		●			●	●	●	●			
	B61		●		●				●		●	
	B62		●		●	●		●	●	●		
DRB1*	0101	●	●			●	●	●	●	●		
DRB1*	0301	●	●	●		●						
DRB1*	0401	●	●	●		●		●				

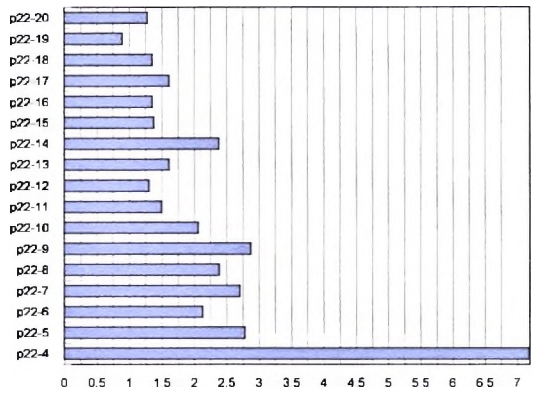
Table 23. Proliferation assay responses of PBMCs to overlapping synthetic HBV core peptide analogues. Red dots represent a stimulation index (SI) of >= 2.5

P22-	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
aa spanned	1-23	4-33	14-43	24-53	34-63	44-73	54-83	64-93	74-103	84-103	94-113	104-123	114-133	124-153	134-163	144-173	154-183
RF1 aa number	635-657	638-667	648-677	658-687	668-697	678-707	688-717	698-727	708-737	718-737	728-747	738-757	748-767	758-777	768-787	778-797	788-817
ES06	●	●		●		●											
ES15	●					●				●			●	●	●		●
ES16						●								●			
HB05		●	●	●			●				●			●			
N231	●																
N232														●	●		
N234											●						
N235																	
N236																	
N237																	
N238																	

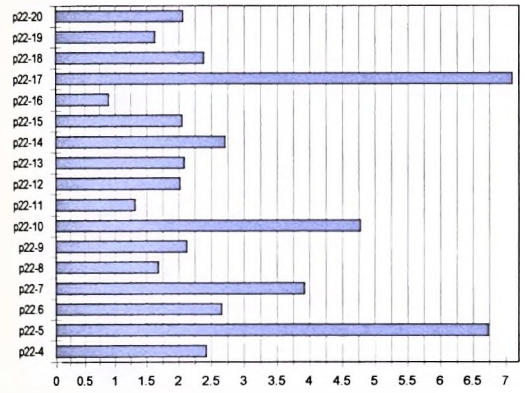
from the four acute and 7 chronic patients are shown in detail in Table 23 and selected graphs are shown in Figure 53. PBMCs from all four acute patients demonstrated a positive proliferative response to between two and 7 of the core peptides. Two of the 7 chronic carriers demonstrated a positive proliferation response to one core peptide while a third carrier showed a proliferative response to two peptides. Four patients (N235, N236, N237 and N238), all of whom were chronic carriers, showed no response to any of the peptides. Published and predicted epitopes, and their position in relation to the 17 overlapping core peptides, are shown in Figure 54.

Three of the 11 patients elicited some reponse to p22-4. ES15 and N231 had moderate

(a)

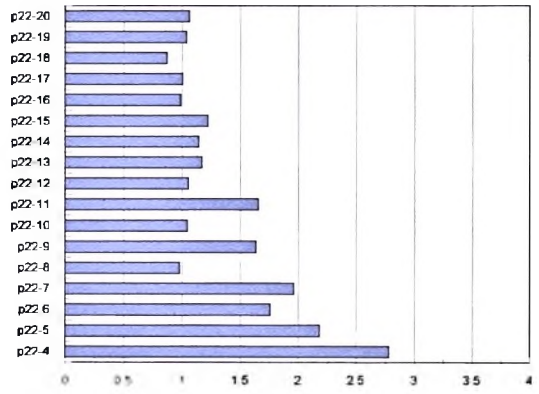


(i) ES06

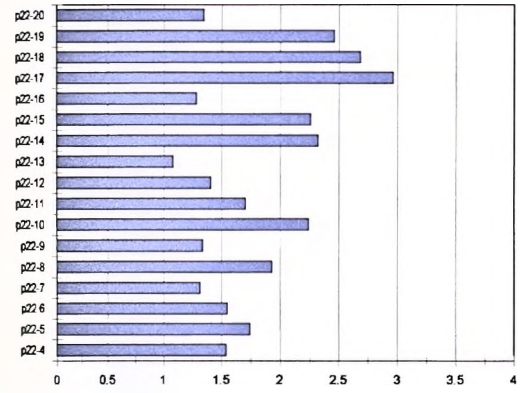


(ii) HB05

(b)



(i) N231



(ii) N232

Figure 53. Stimulation index (SI) plotted against peptide used to prime PBMCs from representative patients with (a) acute hepatitis B (i-ii) and (b) chronic hepatitis B (i-ii)

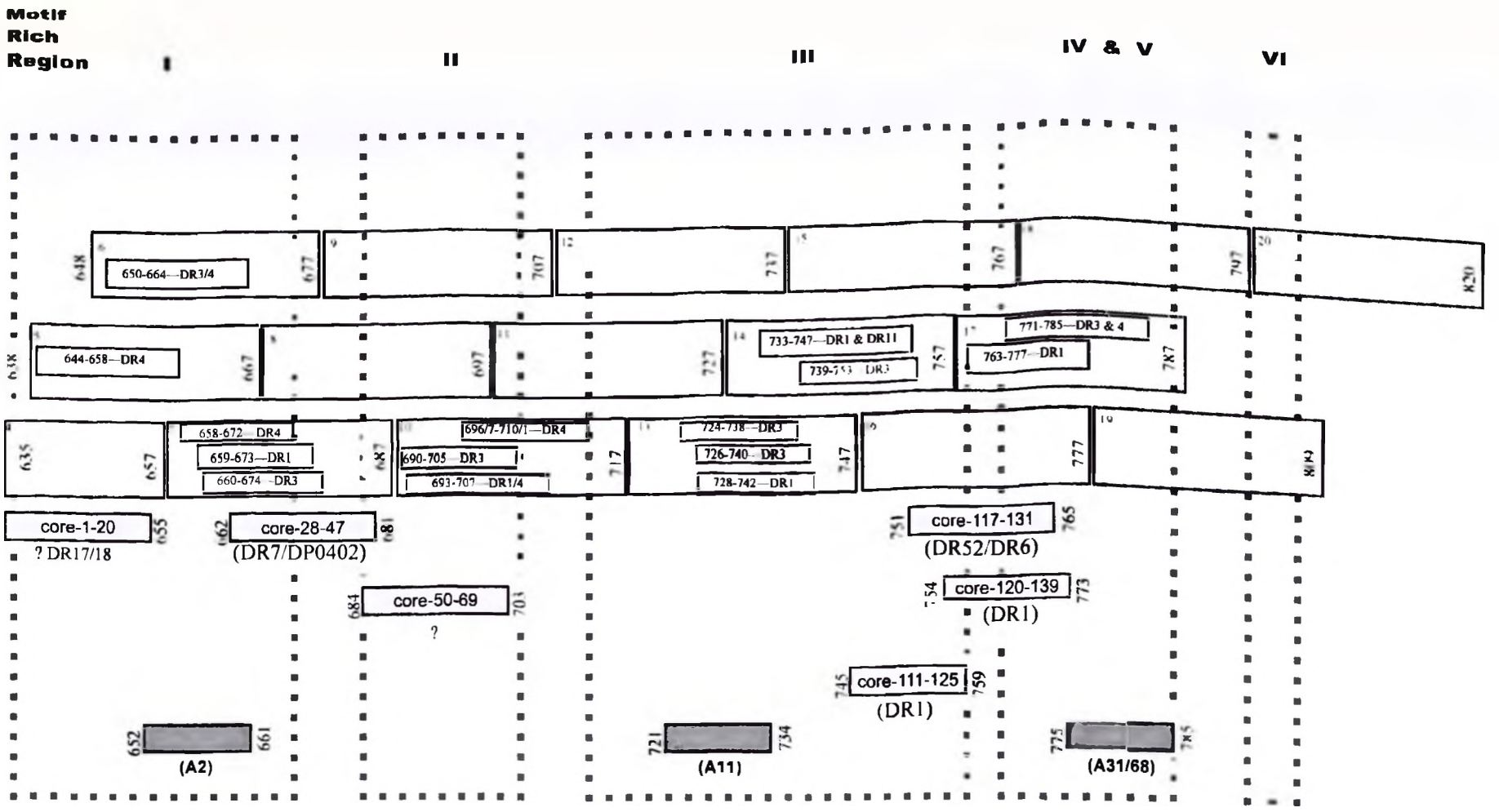


Figure 54. Published class I (dark grey) and class II (light grey) epitopes and their position within motif-rich regions (boxes with dotted borders). The white numbered boxes in the top half of the page represent the overlapping peptides. Smaller boxes within the peptides show the position and specificity of DR motifs found by SYFPEITHI

responses to this peptide whereas ES06 had a strong immunodominant response. Patients ES06 and N231 share the DRB1*0302 allele (DR3, DR18 split) while ES15 is the only specimen typed with the DR17 split of DR3. The p22-4 peptide extends from the start of the core to RF1 amino acid 657 and contains the published DR-specific (allele unspecified) epitope core-1-20.

The two specimens with the DR53 haplotype, ES06 and HB05, were responsive to peptide p22-5. Both these specimens have the DR4 specificity. A DR4 binding motif GATVELLSFLPSDFF (644-658) predicted by SYFPEITHI is found entirely only within p22-5. Underlined amino acids in this and other peptides are suggested anchor residues. A highly ranked DR3– and DR4-specific epitope at RF1 amino acid 650-664 (LSFLPSDFFPSVRDL) is contained completely within both p22-5 (beginning at position 12) and p22-6 (beginning at position 3). This could have been the epitope within p22-6 to which only HB05 responded since a stretch of 9 amino acids, on average starting at absolute positions 3 to 5 of natural ligands, represents the first hydrophobic anchor, or P1, for all known class II ligand motifs (Rammensee *et al.*, 1995).

Both DR4-typed specimens, ES06 and HB05, also elicited a reponse to peptide p22-7 and a DR1, DR4 and DR11 specific epitope (FPSVRDLLDTASALY, RF1 amino acid 658-672) was predicted within the first 15 amino acids of the peptide. Nested motifs PSVRDLLDTASALYR, SVRDLLDTASALYRE and VRDLLDTASALYREA are also predicted in this region by SYFPEITHI and these bind with top 10 DR1, DR3 and DR3, specific epitopes, respectively. These motifs are also all present in p22-6 but at less ideal positions 11-14 of the peptide. LDTASALYREALESP is DR4 specific and only completely contained within p22-7 but starts from position 8 of the peptide. None of the specimens responded to peptide p22-8 (RF1 668-697) and no published or predicted class II epitopes were contained completely in this region which lies between two DR motif-rich areas.

p22-9 extends from RF1 amino acid 678-707 and is the only peptide to contain the entire class II-restricted, published epitope core-50-69 (RF1 amino acid 684-703). Three of the four acute patients (ES06, ES15 and ES16) had a response to peptide p22-9 and one, HB05, elicited a response to peptide p22-10. Although no patient reacts with both peptides, two nested ranked DR3 specific epitopes (RQAILCWGELMTLAT and

QAILCWGELMTLATW) are present within both peptide p22-9 and p22-10. The lower ranked DR1 specific epitope ILCWGELMTLATWVG which is also the first ranked DR4 predicted epitope is also shared by the two peptides. Several DR4 specific peptides (ILCWGELMTLATWVG, WGELMTLATWVGSNL and GELMTLATWVGSNLRE) are present in p22-10 alone.

No responses were recorded for p22-11 and p22-12. Only one specimen, the acute specimen, ES15, had a response to peptide p22-13. p22-13 contains the highly ranked DR1 specific predicted epitope GLKIRQLLWFHISCL and the DR3 specific epitopes NTNMGLKIRQLLWFH and NMGLKIRQLLWFHIS.

One acute patient (HB05) and one chronic carrier (N234) responded to peptide p22-14. The latter specimen was typed DR11. The DR1 specific potential epitope QLLWFHISCLTFGRE is also the seventh ranked DRB1*1101 potential epitope listed by SYFPEITHI (this allele was a recent addition to the list of alleles searched by the SYFPEITHI algorithm). Two potential DR3 specific epitopes (IRQLLWFHISCLTFG and ISCLTFGRETVIEYL) are within p22-14, the first is unique to this peptide while the second is also present within p22-15. Peptides p22-15 (no responders among the patients) and p22-16 (ES15 responded) both contain two published overlapping, DR-restricted epitopes which are DR1- and DR6-restricted, respectively, so would not be expected to cause a response in the patients none of whom were thus typed. p22-16 contains the DRB1*0301 seventh ranked potential epitope SFGVWIRTPPAYRPP. Three of the four acute patients (ES15, ES16 and HB05) and one chronic carrier (N232) responded to peptide p22-17 making this the most immunogenic peptide in my study. p22-17 shares the predicted DR1 restricted epitope (PPAYRPPNAPILSTL) with p22-16. APILSTLPETTVVRR is both DR3 and DR4 specific and is contained in both p22-17 and p22-18. Two of the specimens, ES15 and N232, reacted to both p22-17 and p22-18.

There are no published class II epitopes in the 3' end of the core which was represented in the last three peptides p22-18, p22-19 and p22-20 and no DR favoured motifs were found by SYFPEITHI beyond the DR3 restricted binding site contained within p22-18 at RF1 781-794. However a motif scanner at http://phage.lanl.gov/cgi-bin/EPI_PREDICT/MultiMotifScanner.pl which scans for a simpler but wider range of DR motifs predicts

several DR specific motifs at the end of the molecule, including three specific for the DRB3 alleles of the DR52 haplotype. Only one patient, HB05, of the many patients with the DR52 haplotype, had a moderate response to p22-20.

9.4 Discussion

All of the acute patients responded to more than one region of the core while only one chronic carrier, N232, had a response to more than one peptide and only three of the chronic carriers responded at all. This is in agreement with the finding that chronic carriers have weak and more narrowly focussed T-cell responses while patients who clear the disease have more vigorous, multispecific responses (Chisari and Ferrari, 1995). Very weak proliferation responses are routinely found in asymptomatic, nonviraemic HBsAg-positive carriers with normal aminotransferase levels (Lohr *et al.*, 1995) and since 90 percent of adult South African chronic carriers fit this profile (Dusheiko *et al.*, 1985) I anticipated low SI values, particularly in the chronic carriers. I was, however, unable to obtain specimens with acute HBV from Namibia and reportedly acute hepatitis B is a very rare disease outcome in Namibia (K. Seidel, personal communication). All 7 of the DR-typed chronic carriers had seroconverted to anti-HBe which is an indication that they had indeed progressed to predominantly nonreplicative infection. The multi-faceted response I observed in the patients with acute hepatitis would mean that in these patients selection of a strain mutated at an anchor residue would probably not abrogate the immune response, but merely shift the immunodominant response to another epitope.

The core gene has both class I and class II binding motif-rich regions. Epitopes were predicted between RF1 amino acids 638-672, 684-705, 709-761, 765-774, 778-785 and 798-802 and published epitopes were largely found within these motif-rich regions. The 12 amino acid insert present only in genotype G specimens introduces several additional T-cell motifs to the first part of the core gene product. Genotype G sequence data is limited and the few sequences available on GenBank do not show variation within this insert.

The first predicted epitope which is covered by peptides p22-4 to p22-7 is the most motif-rich region of the core. The two DR18 typed specimens, ES06 and N231, have an immunodominant response to p22-4. ES15 which was typed DR17 also had a positive, but not an immunodominant response, to this peptide. No favoured DR motif was found by

several DR specific motifs at the end of the molecule, including three specific for the DRB3 alleles of the DR52 haplotype. Only one patient, HB05, of the many patients with the DR52 haplotype, had a moderate response to p22-20.

9.4 Discussion

All of the acute patients responded to more than one region of the core while only one chronic carrier, N232, had a response to more than one peptide and only three of the chronic carriers responded at all. This is in agreement with the finding that chronic carriers have weak and more narrowly focussed T-cell responses while patients who clear the disease have more vigorous, multispecific responses (Chisari and Ferrari, 1995). Very weak proliferation responses are routinely found in asymptomatic, nonviraemic HBsAg-positive carriers with normal aminotransferase levels (Lohr *et al.*, 1995) and since 90 percent of adult South African chronic carriers fit this profile (Dusheiko *et al.*, 1985) I anticipated low SI values, particularly in the chronic carriers. I was, however, unable to obtain specimens with acute HBV from Namibia and reportedly acute hepatitis B is a very rare disease outcome in Namibia (K. Seidel, personal communication). All 7 of the DR-typed chronic carriers had seroconverted to anti-HBe which is an indication that they had indeed progressed to predominantly nonreplicative infection. The multi-faceted response I observed in the patients with acute hepatitis would mean that in these patients selection of a strain mutated at an anchor residue would probably not abrogate the immune response, but merely shift the immunodominant response to another epitope.

The core gene has both class I and class II binding motif-rich regions. Epitopes were predicted between RF1 amino acids 638-672, 684-705, 709-761, 765-774, 778-785 and 798-802 and published epitopes were largely found within these motif-rich regions. The 12 amino acid insert present only in genotype G specimens introduces several additional T-cell motifs to the first part of the core gene product. Genotype G sequence data is limited and the few sequences available on GenBank do not show variation within this insert.

The first predicted epitope which is covered by peptides p22-4 to p22-7 is the most motif-rich region of the core. The two DR18 typed specimens, ES06 and N231, have an immunodominant response to p22-4. ES15 which was typed DR17 also had a positive, but not an immunodominant response, to this peptide. No favoured DR motif was found by

SYFPEITHI in p22-4 although the peptide contains the published DR specific epitope, core-1-20 which EPI_PREDICT lists as being a potential epitope of DR17. DR17 favours an Asp in pocket four (Rammensee *et al.*, 1995) and an aromatic amino acid in pocket 9 making MDIDPYKEF a potential epitope for the two splits of DR3, DR17 and DR18.

The two DR4 typed specimens, ES06 and HB05, share a positive response to both p22-5 (core-4-33) and p22-7 (core-24-53) both of which contain SYFPEITHI predicted DR4 binding motifs. Class II epitopes have not been reported previously in either of these regions.

p22-8 spans the region between the first and second motif-rich regions and encompasses no predicted or published epitopes and elicited no response from any of the patients. The second HLA motif-rich region (684-705) coincides almost exactly with the published class II epitope which Ferrari *et al.* (1991) found in all but one of their patients (core-50-69) and all of the acute patients reacted to this region which was completely spanned by p22-9 (core-44-73) and partially spanned by p22-10 (core-54-83). The high incidence (95.6 percent) of immunogenicity of this epitope in Ferrari's study compares with an incidence of 69 and 73 percent for two other regions, core-1-20 and core-117-131, respectively Three of the acute patients (ES06, ES15 and ES16) responded to p22-9 while the fourth acute patient (HB05) responded to p22-10.

The third motif-rich region extends from 709-761. The most dense motif frequency extends over a narrow 6 amino acids 713-718 in this region (Figure 52). Mapping the SYFPEITHI predicted class II epitopes shows this to be between two class II motif-rich regions. No published T-cell epitope has been mapped to this region. The region which extends from RF1 amino acid 712 to 723 is present on the surface of the capsid molecule and is a major B cell epitope of HBV. It contains both the single conformational determinant responsible for the HBc antigenicity in the assembled core (HBc) and a linear HBe-related determinant (HBe1, Salfeld *et al.*, 1989). The motif-rich region upstream to this contains a published HLA-A11 restricted epitope (core-87-100, RF1 amino acid 721-734). No significant T-cell responses were recorded by the *in vitro* proliferation assay for either peptide p22-11 (697-727) or p22-12 (707-737).

SYFPEITHI identified two DR17 sites in p22-13 and ES15, the only specimen with this split of the DR3 allele, was also the only specimen that responded to this peptide. The most immunogenic region of the core was found to be contained between RF1 amino acids 758-787. This region is spanned by the p22-17 peptide and contains the fourth and fifth motif-rich regions. The response of HB05 to this region was particularly strong. Three of the four acute patients and one chronic carrier responded to peptide p22-17, and all responders had the broad DR3 allele. The three specimens which did not respond included N234 (DR11) which did have a weak, but below cut-off response, and the two specimens which have the DR18 split of DR3 (ES06 and N231). Peptides p22-15 to p22-17 cover the region of the genome where published reports have found three overlapping DR1 and DR6 specific class II epitopes.

A published class I HLA-A31 or HLA-A68 restricted epitope extends from RF1 amino acid 775-785 (core-141-151) within a region of less dense binding motif concentration towards the 3' end of the core. This class I epitope is within peptide p22-18. Several DR3 and DR4 motifs were found in peptide p22-18 but beyond this peptide SYFPEITHI does not predict many high binding class II epitopes.

Thus there are few predicted and no published class II epitopes in the sixth and last binding motif cluster at the carboxy-terminus of the core gene from RF1 amino acid 798-802. Re-examination of the raw data revealed that this arginine rich part of the core contains many B27- and A3-supertype motifs and it is these which cause the last motif cluster on the histogram. SYFPEITHI found two overlapping potential DR11 (a split of DR2) epitopes. It was interesting that the carboxy-terminus also contained several HLA-A2 binding motifs which had negative binding affinities. Since the carboxy-terminus is well conserved not only within the HBV strains but also within mammalian hepadnaviruses it does not appear to be under considerable selection pressure. Nonetheless, one of my specimens, ES15, which has the allele DR1605, which is rare in Africa, had a proliferative response to peptide p22-20. No information is available regarding the favoured motifs of the DR16 split. The carboxy-terminal region of the core has not previously been reported to be immunogenic.

The finding of different immunodominant target molecules in different populations is to be expected as the frequency of the various HLA alleles is ethnically distributed. The location

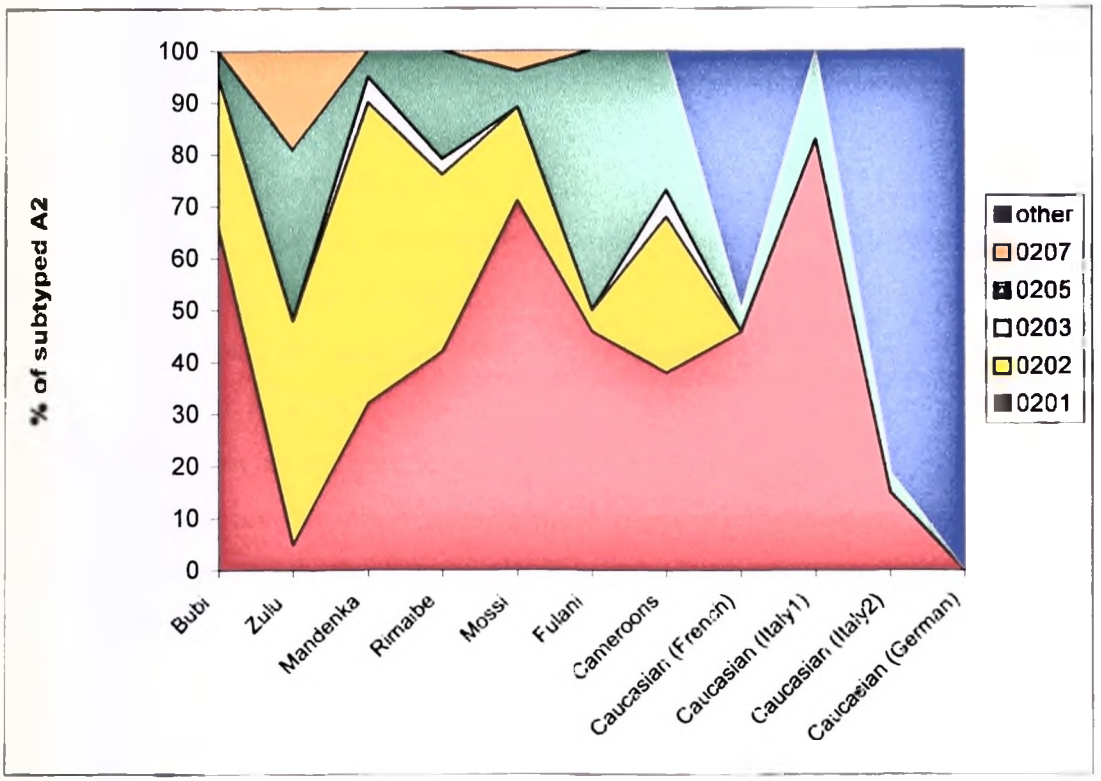


Figure 55. Distribution of A2 subtypes in selected sub-Saharan Africans and Europeans (Table 16 gives details of references used for population frequency data)

of favoured DR motifs was limited by the fact that (at the time of the analysis) SYFPEITHI only searched for anchor motifs of DRB1*0101, DRB1*0301 (i.e. the DR17 split of DR3) and DRB1*0401. DR11 has been recently added to it's repertoire. However, since DR favoured motifs are more promiscuous than class I alleles and tend to cluster together there was a good correlation between DR type, potential epitopes and the responses that were observed. Published motifs also clustered within the regions of high motif density with the exception of three of the DR specific epitopes which overlapped the predicted epitope regions. This may have been due to the lack of favoured motifs for the DR alleles in the SYFPEITHI database.

While the frequency of individuals positive for a given particular allele varies drastically, when the overall frequency of each supertype is calculated coverage of about 35 to 55 percent is predicted irrespective of ethnicity (Sidney *et al.*, 1996; Sette and Sidney, 1999). This is important for the development of peptide therapeutic vaccines (Bertoni *et al.*, 1997) but while cross-reactive, not all alleles bind the same epitopes equally (de Guercio *et al.*, 1995). For example, SYFPEITHI produces a list of 202 possible core epitopes for the HLA-A0201 allele, 196 of these have a binding score greater than one and 88 have a binding

score greater than 10 while the allele HLA-A0202 yields a list of only 10 possible epitopes all of which have a binding score of one. The HLA-A0203 allele yields a list of 41 possible epitopes, 31 of which have a binding score greater than one, but only two have a binding score greater than 10. Comparison of the BIMAS listings for HLA-A0201 and HLA-0205 shows similarly that the former allele binds putative epitopes with much higher scores than the latter. Although the HLA-A0201 allele is common in many populations it is rare or absent in others (Figure 55) and this includes the South African Zulus in whom the most frequent A2 allele is A*0202. If the binding scores calculated by the algorithms of the epitope prediction programmes are related to their immunogenicity as their authors believe, these and other low binding alleles found more frequently in sections of the population could be responsible for the blocking, or antagonistic, effect which has been reported in HBV (Bertoletti *et al.*, 1994) which could be a factor in HBV disease outcome.

CHAPTER 10

HEPATITIS B VIRUS CORE GENE VARIATION WITHIN PREDICTED AND PUBLISHED EPITOPES

10.1 Introduction

The host immune response critically determines the clinical presentation and outcome of HBV-infection (Meuer and Moebius, 1994). A vigorous and polyclonal T helper cell (Th) response to the core antigens is found in most patients with acute disease (Ferrari *et al.*, 1990; Jung *et al.*, 1995). In contrast, chronic persistence is characterized by less-vigorous Th responses, a low frequency of HBV-specific Th cells in the peripheral blood and in the liver and weak or undetectable virus-specific CTL responses. The HBcAg and HBeAg have been shown to be cross-reactive at the T-cell level (Milich *et al.*, 1987) and numerous class I and class II binding sites have been defined and identified (Chapter 9). Conservative amino acid change where the amino acid replacement is closely related both chemically and physically to its alternate, will result in little change in the binding affinity of an immunogenic peptide. However, the converse, where mutations result in non-conservative change, can lead to immune escape. To determine whether changes within the virus could have an influence on the immune response and thus the outcome of the disease locally, I sequenced the core gene of isolates from chronic carriers from southern Africa and compared this sequence data with 161 global core gene sequences from GenBank.

10.2 Materials and Methods

10.2.1 Patients

Fourteen patients, five from South Africa and 9 from Namibia, were chosen for core gene sequencing. All of the Namibians (JB01, 02, 04, 05, 07, 11, 13, 16 and 17) and one of the South Africans (ES05) were infected with HBV of genotype E. Two of the South Africans were infected with genotype C (ES01 and ES03) and the remaining two were infected with genotype D (ES04) and A' (ES02), respectively. The core sequence of these local specimens was compared with 161 core genes from GenBank.

10.2.2 Sequencing

DNA was extracted as described in paragraph 3.2.1. The full core gene was amplified as described in paragraph 3.2.2 using sense primer Sn1738 and antisense primer Asn2467 to give an amplicon with a size, dependent on genotype, of 748-754 base bairs. PCR fragments were sequenced manually as described in paragraph 3.2.4.

10.2.3 Phylogenetic analysis

A neighbour-joining tree showing the relationships between the 175 core genes was drawn using software from the PHYLIP package (version 3.573c, Felsenstein 1993). The 175 specimens included 16 subgroup A, four subgroup A', 31 genotype B, 68 genotype C, 35 genotype D, 12 genotype E, three genotype F, two genotype G and four chimpanzee HBV sequences.

The nucleotide and protein "hotspots" within the core database were calculated as described previously (Chapter 6). The recombinant identification programme was re-run to determine whether any further putative recombinants could be identified among the additional specimens. The precore/core sequence data (including the genotype G 12 nucleotide insertion) was also partitioned as before into 11 files of 60 nucleotides each (and one file of 21 nucleotides only), and the relationships between the genotypes determined and plotted for each of the 12 intervals as described (Chapter 7).

10.3 Results

10.3.1 Phylogenetic analysis

The core neighbour-joining tree is shown in Figure 56. Genotype B was extended from 12 to 31 specimens for this study and 14 of these additional specimens clustered with the original subgroup B' specimens (B'01-B'04) in the genotype C clade. The remaining 5 additional specimens clustered with the original 6 subgroup B of genotype B specimens which partition independently in all subgenomic regions. The third subgroup, B", retained its two original specimens but none of the additional specimens clustered in this clade.

Core sequence data from 10 southern Africans found to be genotype E increased the number of genotype E specimens to 12 for this analysis. The additional specimens, however, made

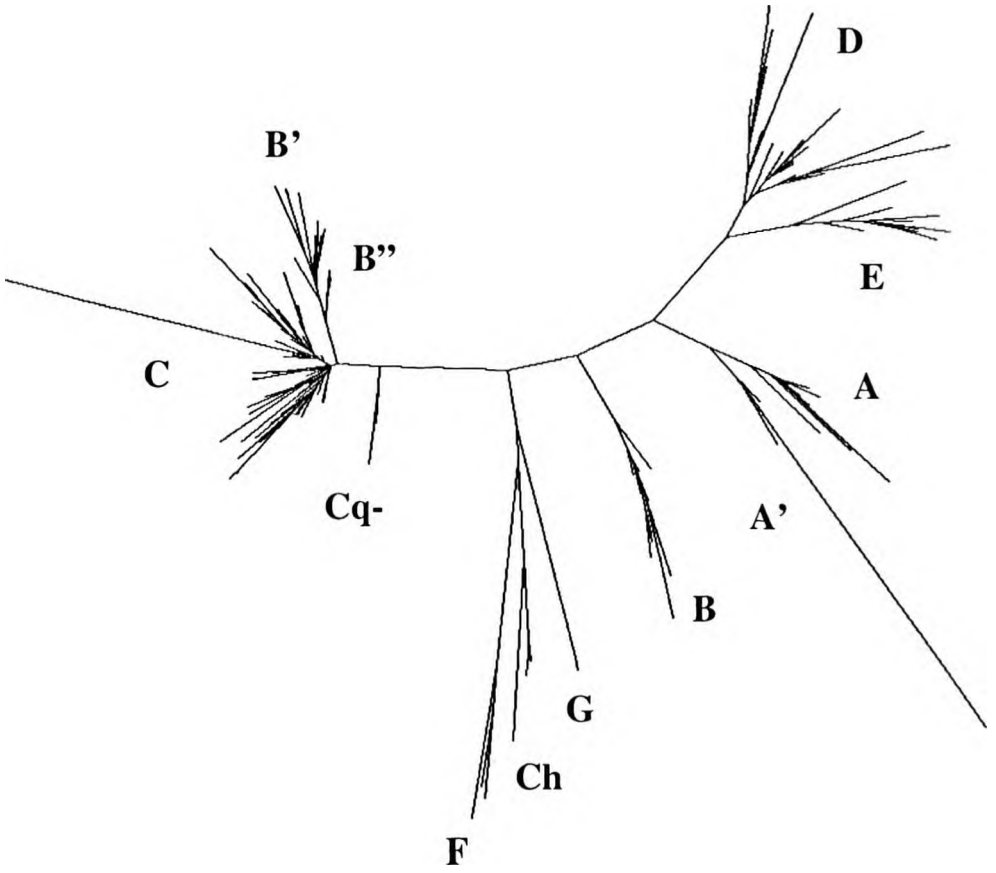


Figure 56. Neighbor tree (UPGMA option) of 175 core gene sequences
The tree was drawn using the unweighted pair-group method with arithmetic mean (UPGMA) option of the PHYLIP NEIGHBOR programme. This is an average distance method that handles multiple sets of data easily and quickly making it suitable for bootstrap re-sampling studies. The tree is constructed by successive (agglomerative) clustering using an average linkage method of clustering and unlike neighbor joining the UPGMA option gives a mid-point rooted tree (Felsenstein 1993)

little difference to the clustering of the genotype E clade which partitions as a subgroup of genotype D in the core gene region.

Figure 57 shows the output of the DNADIST matrix averaged to show the mean difference between the clades in the tree. The difference between genotypes over the core gene varies from 8 to 14 percent, subgroups differ by 4 to 8 percent and specimens within a subgroup differ by less than 4 percent. Clearly, subgroups B' and B" of genotype B form a subgroup of genotype C, genotype E forms a subgroup of genotype D and specimen D18 groups with genotype A in this region. Genotype D should not be partitioned into subgroups and genotype F should be partitioned into subgroups on the basis of these cut-off values. The one exception to the intergroup definition is in subgroup A' where the heterogeneity of A'02, caused by a frameshift, raises the non-definitive variation within the subgroup to more than 6 percent.

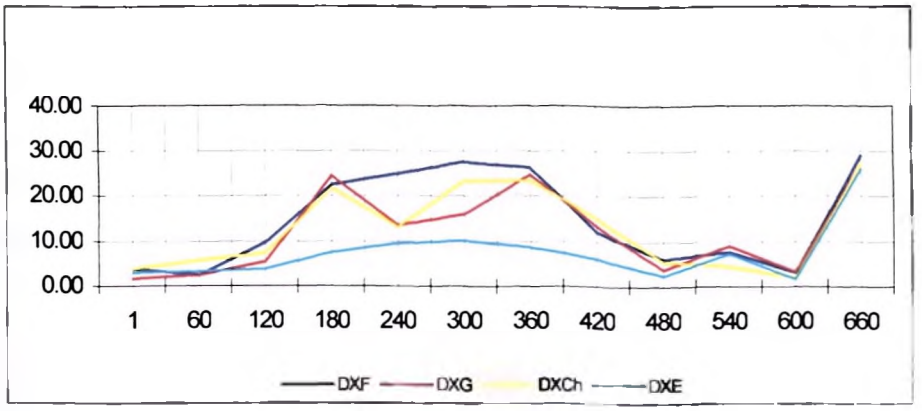
	D18	A	A'	B'	B	B"	C	Cq-	D	Dm	E	F	G	Ch
D18	0.00													
A	3.55	1.72												
A'	7.16	6.03	6.24											
B'	9.84	10.42	12.27	2.11										
B	8.64	8.72	9.71	9.31	3.04									
B"	9.64	9.94	12.00	2.92	9.68	1.11								
C	10.27	10.47	12.16	3.90	9.81	3.47	2.89							
Cq-	10.31	10.52	11.67	4.88	9.28	4.62	4.14	2.39						
D	7.15	8.47	9.75	10.12	9.48	10.00	10.26	10.15	4.01					
Dm	5.99	7.50	9.15	9.65	9.16	9.52	13.61	9.26	3.84	1.44				
E	7.87	8.52	9.91	11.11	10.14	10.95	11.30	10.93	6.37	5.60	2.90			
F	13.02	12.99	13.99	11.57	12.38	11.61	11.24	10.20	13.57	12.46	13.12	5.04		
G	12.35	11.25	12.94	10.81	11.34	11.02	11.16	10.25	11.52	10.42	10.52	11.47	0.30	
Ch	11.35	10.99	11.91	10.57	10.18	10.15	9.94	9.14	11.52	11.43	10.72	10.91	10.06	2.90

Figure 57. Average nucleotide differences between specimens (N = 175) from genotypes A to G and an additional group, Ch, which contains four specimens from the newly identified chimpanzee hepatitis B virus. Boxes highlight values outside the normal inter- and intra-genotype and subgroup ranges (which can be the result of recombination) as described in the text and investigated in detail in Chapter 8. A red border highlights inter-genotype distances less than the 8 percent cutoff, blue indicates inter-subgroup distances greater than 8 percent and green indicates intra-subgroup distances greater than 4 percent

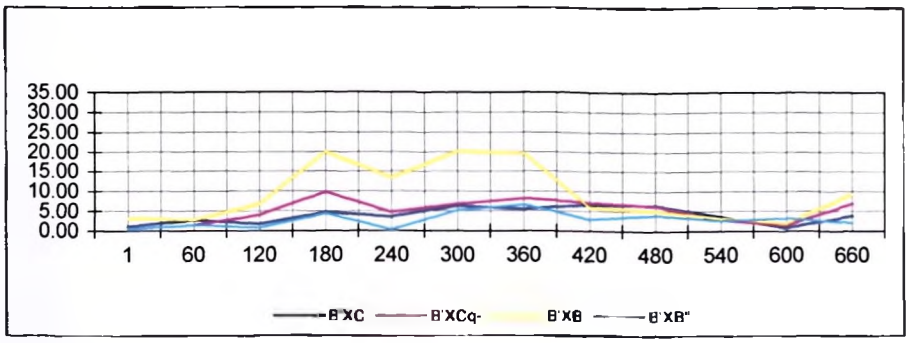
10.3.2 Core type-relationship graphs

The core type-relationship graphs (Figure 58) show that the first three intervals (or 180 nucleotides) and the last three full intervals of the core gene are very well conserved. Between the two conserved regions is a heterogeneous middle section with differences of 20 percent or more. The last interval, which consisted of only 21 nucleotides, is alternatively well conserved (Figure 58b and c) or heterogenous between different subgroups (Figure 58a), partitioning genotypes A to D together and away from genotypes E to G and chimpanzee hepatitis B. All members of the subgroups B' and B" which cluster away from the third subgroup of genotype B in the core gene have genotype differences with subgroup B, but non-definitive differences with the two subgroups of genotype C (Figures 57 and 58b).

(a)



(b)



(c)

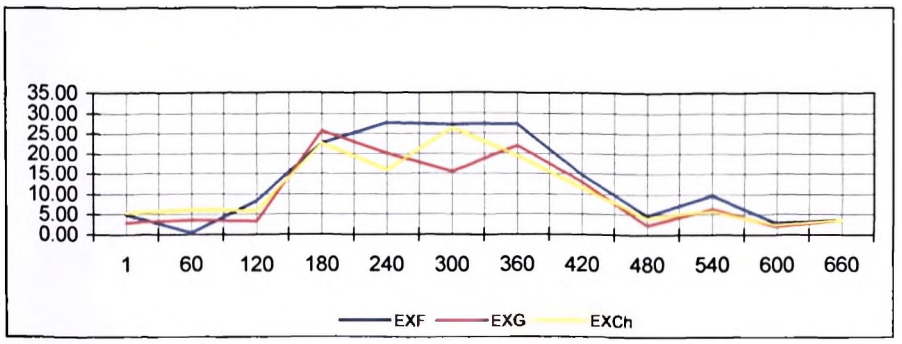


Figure 58. Average genetic distance between the genotypes specified plotted against the first nucleotide number of the interval (with 1 representing the start of the preCore) over the 645 nucleotides of the preCore/core gene. In all graphs the core gene can be seen to be made up of two conserved regions sandwiched between a more variable region. (a) Genotypes D and E have only intra-genotypic differences over most of the core gene, (b) subgroup B' of genotype B is more closely related to the two subgroups of genotype C and subgroup B" than it is to the third subgroup of genotype B, subgroup B, and (c) only normal inter-genotypic differences were found between genotype E and genotype F (blue) and genotype E and the new genotypes, genotype G (red) and the chimpanzee hepatitis B virus (yellow)

Similarly genotype E forms a subgroup of genotype D over this region (Figures 57 and 58a). The additional South African genotype E specimens do not differ markedly from the two prototype genotype E specimens in the core protein. Only normal inter-genotypic differences were found between genotype E, genotype F and the new genotypes, genotype G and the chimpanzee hepatitis B virus (Figure 58c).

10.3.3 Variation within the predicted and published epitopes

Seven regions were predicted to contain the T-cell linear epitopes of HBV. These included the genotype G 12 amino acid insert and 6 motif-rich regions (Figure 52), labelled I to VI in Figure 54 and occurring between the following RF1 amino acids: 635-672, 684-705, 709-761, 765-774, 778-785 and 798-802 (Chapter 9).

10.3.3.1 The genotype G insert

The 12 amino acid insert at the start of the core gene in genotype G was found to introduce many additional MHC binding motifs. The most important anchor residues within potential peptides from this insert are the arginine at position one, the leucines at position 8 and 11 and the proline at position five (Figure 59). No variation has been reported in the insert of genotype G specimens studied to date.

10.3.3.2 Potential T-cell epitope region I (amino acid 635-672)

The motif-rich region of the core extending from RF1 amino acid 635-672 is the most

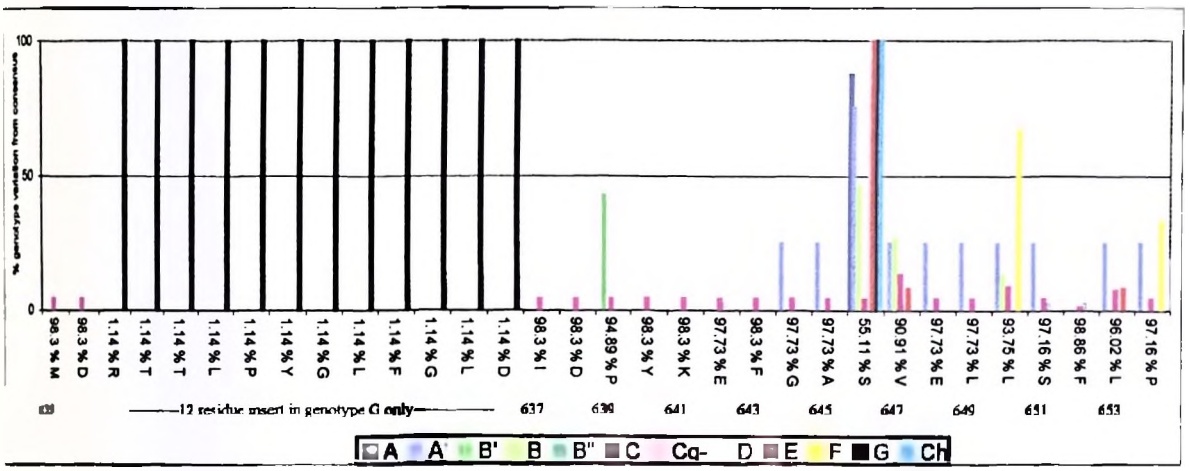


Figure 59. Variation within published epitope core-1-20. The twelve amino acid insert definitive of genotype G is within this epitope. For this and the following six graphs, the consensus amino acid and its frequency is shown along the x axis and colour coded bars show percentage variation within each genotype at each position

motif-dense region of the genome. Three published epitopes have been reported within this region (Figures 59-61). The first of these covers the first 20 amino acids of the core gene, RF1 amino acid 635-654 (core-1-20, Figure 59), and is reported to be HLA class II restricted (Penna *et al.*, 1991).

The most widespread variation in the first published epitope is the serine to threonine switch at position 646 which is genotype specific (Figure 59). This switch retains moderate conservation of amino acid type and properties (Taylor, 1986) and SYFPEITHI does not predict a loss of class II binding properties as a result of the ^{646}Ser → ^{646}Thr switch but several of the predicted class I epitopes have higher binding scores with ^{646}Thr. The less common proline to threonine switch at amino acid 639 on the other hand is found to abrogate the principal epitope (DPYKEFGAS) presented to, amongst others, the HLA-B5101 allele. ^{647}Val exhibits several different amino acid changes. One such change is a valine to glutamic acid switch which changes the binding properties of several predicted epitopes (e.g. the SYFPEITHI binding value increases from five for the epitope ASVELLSFL to 15 for ASEELLSFL). ^{650}Leu undergoes several conservative non-definitive mutations (B16, B17 and C68 switch to isoleucine while C16 and C17 switch to valine) but two thirds of genotype F specimens have a poorly conserved leucine to phenylalanine switch in the last anchor residue which lowers the binding value from 11 for EFGASVELL to one for EFGASVELF.

The second published epitope within this region is short and class I restricted and reported

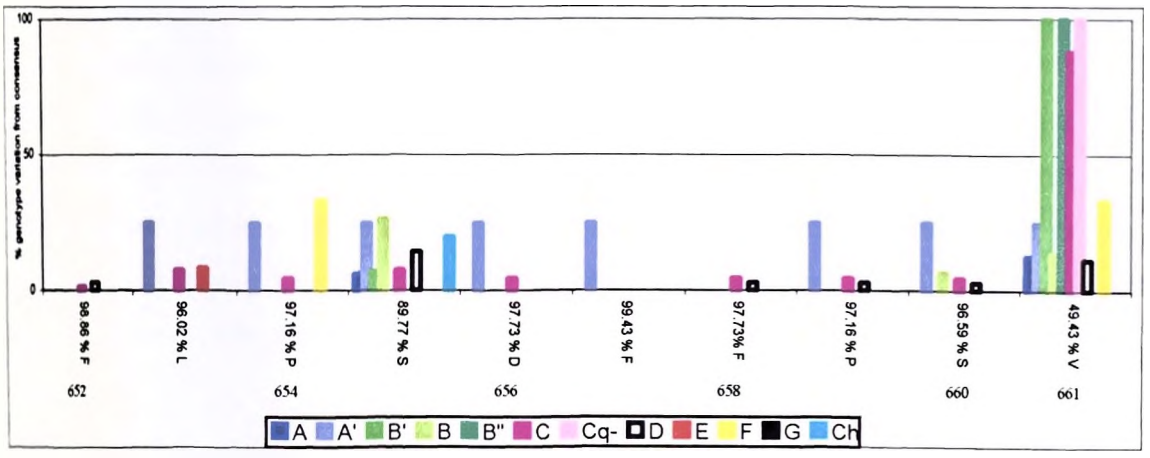


Figure 60. Variation within the published class I HLA-A2 restricted epitope core-18-27

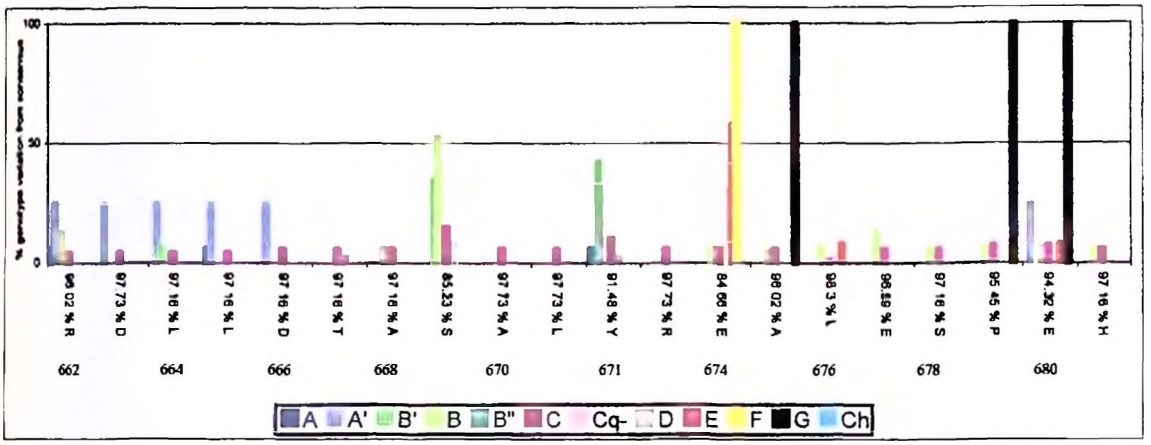


Figure 61. Variation within published epitope core-28-47

to be specific for the HLA-A2 allele. It extends from amino acid 652-661 (core-18-27, Figure 60). The overall motif density of this region is less than for the preceeding and subsequent regions. The most common variation in this region is a highly conservative valine to isoleucine change at RF1 amino acid 661. Although the change to isoleucine is moderate in amino acid type and properties, both changes lower the BIMAS score by a factor of 10. This change is subgroup B versus subgroups B', B", and C genotype-specific and by lowering binding it could alter or abrogate the immunogenicity of the epitope. ^{655}Ser undergoes fairly common non-definitive change and this is predicted to have a variable effect on binding. For example, a serine to proline change in the auxilliary residue of epitope PSDFFPSVR changes the SYFPEITHI Score for the A1 restricted epitope from a value of 15 to one of 11 in the peptide PPDFFPSVR.

The last published epitope of this region extends from RF1 amino acid 662-681 (core-28-47, Figure 61) stretching beyond the region of most dense motif clustering and is reported to be class II restricted, specific for the DR7 and DP0402 alleles in particular. A definitive amino acid change at position 674 is very well conserved (aspartic to glutamic acid) and there is a common non-definitive change at position 669. Subgroup B varies from the rest of genotype B at this position having an alanine instead of a threonine or serine. The BIMAS score of the HLA-B7 allele SVRDLLDTS rises from 1 to 5 when the serine anchor residue is exchanged thus.

10.3.3.3 Potential T-cell epitope region II (amino acid 684-705)

HLA motif-dense region II extends from RF1 amino acid 684-705 (core-50-69) and

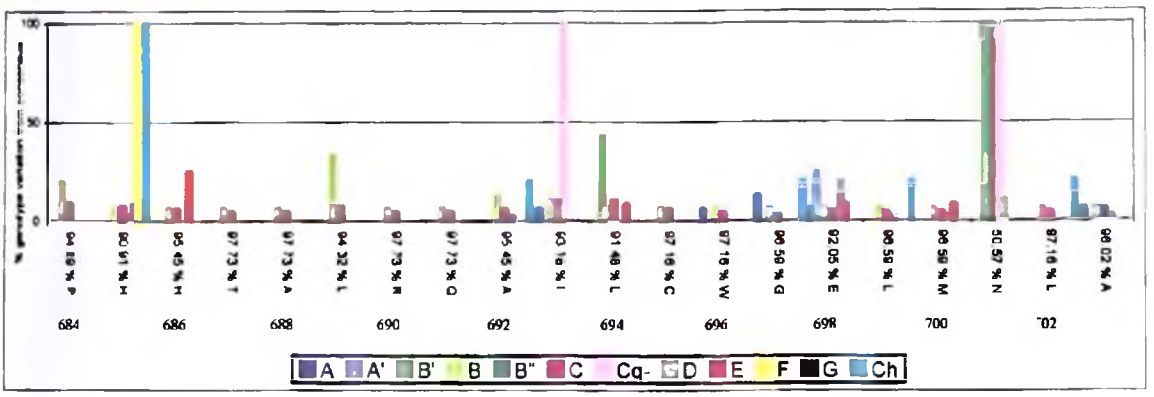


Figure 62. Variation within the published class II restricted epitope core-50-69

coincides almost exactly with the published epitope from RF1 amino acid 684-703 (core-50-69) which Ferrari *et al.* (1991) found in all but one of their patients (Figure 62). The high incidence (95.6 percent) of immunogenicity of this epitope in their study compares with an incidence of 69 and 73 percent for two other regions, core-1-20 and core-117-131, respectively.

The commonest change in the published epitope is a threonine to asparagine change at amino acid 701. This switch is genotype specific, being present in the two genotype B subgroups B' and B" and in genotype C. This change had some effect on binding values but did not abrogate any of the epitopes tested. Other definitive changes include the conservative isoleucine to valine switch at RF1 amino acid 693 which separates subgroup Cq⁻ from the rest of genotype C and the histidine to asparagine change at 685 which is definitive of genotype F. There are several other minor positions of non-definitive change in the published epitope (Figure 62). These range from moderately conserved change at RF1 amino acid 684 (proline to histidine, serine or adenine) and 701 to well conserved change at 689, 694, and 698.

10.3.3.4 Potential T cell epitope region III (amino acid 709-761)

The first peak in the third motif-dense region extends from 709-720 but the most dense motif frequency extends over a narrow 6 amino acid (713-718) region. No published T-cell epitope is entirely contained within this region but it is an important, surface exposed, B cell epitope (the e1 epitope) of both HBeAg and HBcAg. There is weak to very weak conservation of amino acid type and properties at amino acids 708, 713, 715 and 718. There is well conserved definitive and non-definitive change (the consensus for subgroup B is

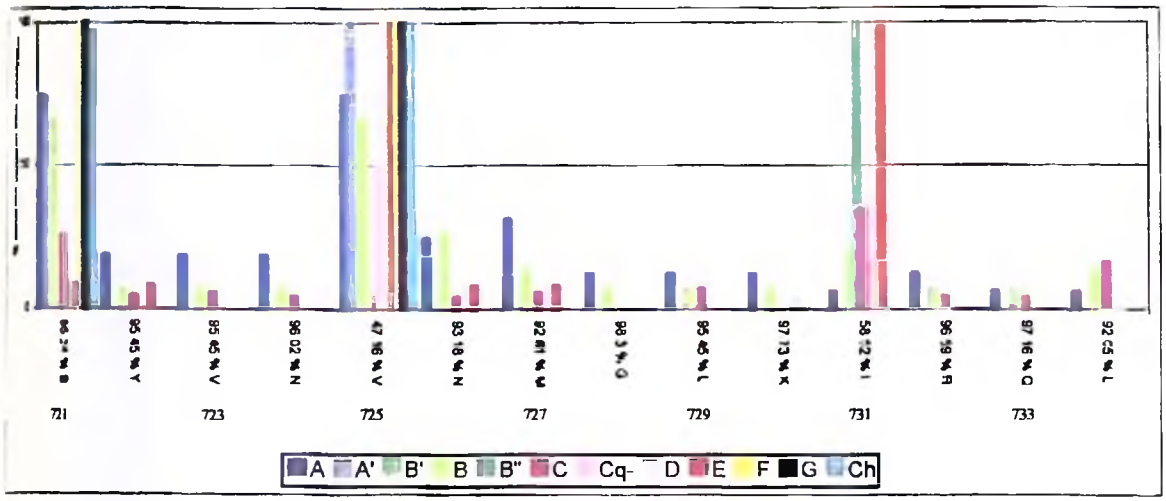


Figure 63. Variation within the published class I HLA-A11-restricted epitope core-87-100

glutamine where all other groups and subgroups are glutamic acid) at amino acid 711 and moderate non-definitive change at 714. This region is also commonly deleted in core deletion mutants. The latter are particularly frequent (20 percent) among the genotype A specimens of our core database (A10, A11 and A12).

The second T-cell motif-rich region within the third predicted epitope extends from RF1 amino acid 726-738 and partially contains a published HLA-A11 restricted epitope (core-87-100, RF1 amino acid 721-734, Figure 63). BIMAS only searches for the binding motifs of the HLA-A*1101 allele, but 8 of the top ten core epitopes of this allele are within the region 719-746 albeit with low scores ranging from 0.012-2.0.

There are moderately conserved definitive changes at RF1 amino acids 721 (serine to asparagine) and 725 (valine to threonine) of this epitope. Both of these changes differentiate between subgroup B and the two recombinant subgroups of B, B' and B".

A frequent mutation at position 731 (isoleucine to leucine or phenylalanine) was recently reported to be responsible for an "immature secretion" phenotype which results in the excessive secretion of virions containing an immature viral genome (Yuan and Shih, 2000). Sixty-four (36.6 percent) of the specimens in the core database had this variation. The phenylalanine variant is definitive to genotypes D and E while the leucine variant is a common inter-subgroup variation occuring in 27.9 percent of genotype C specimens. This abnormality is offset by another frequent core mutation (proline to threonine or glutamine)

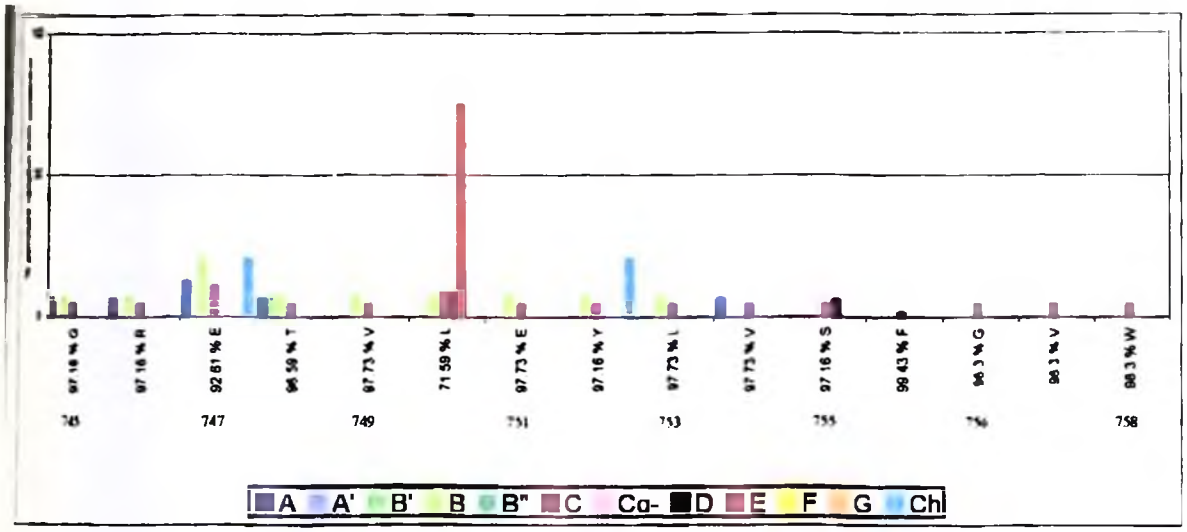


Figure 64. Variation within the published class II DR1-restricted epitope core-111-125

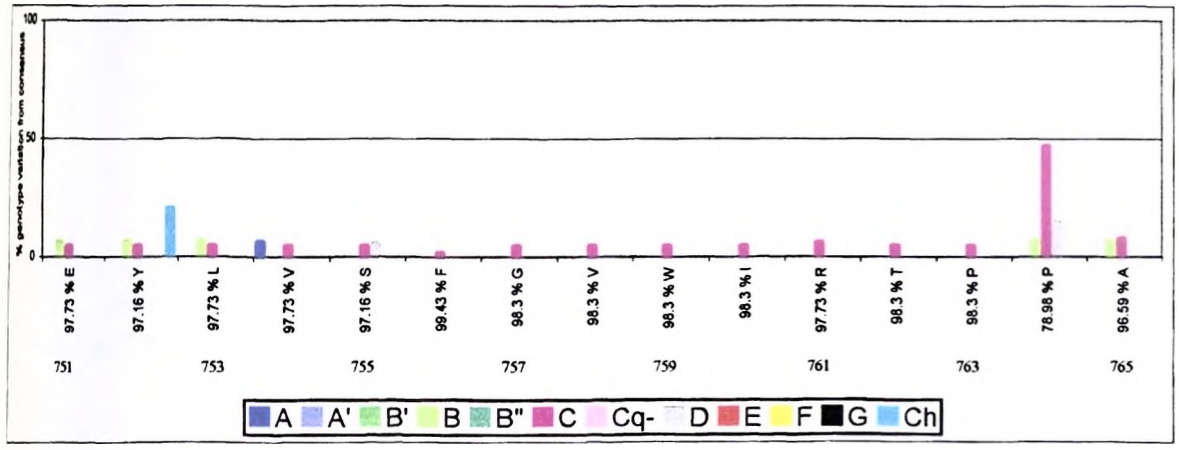


Figure 65. Variation within the published class II DR52/DR6-restricted epitope core-117-131

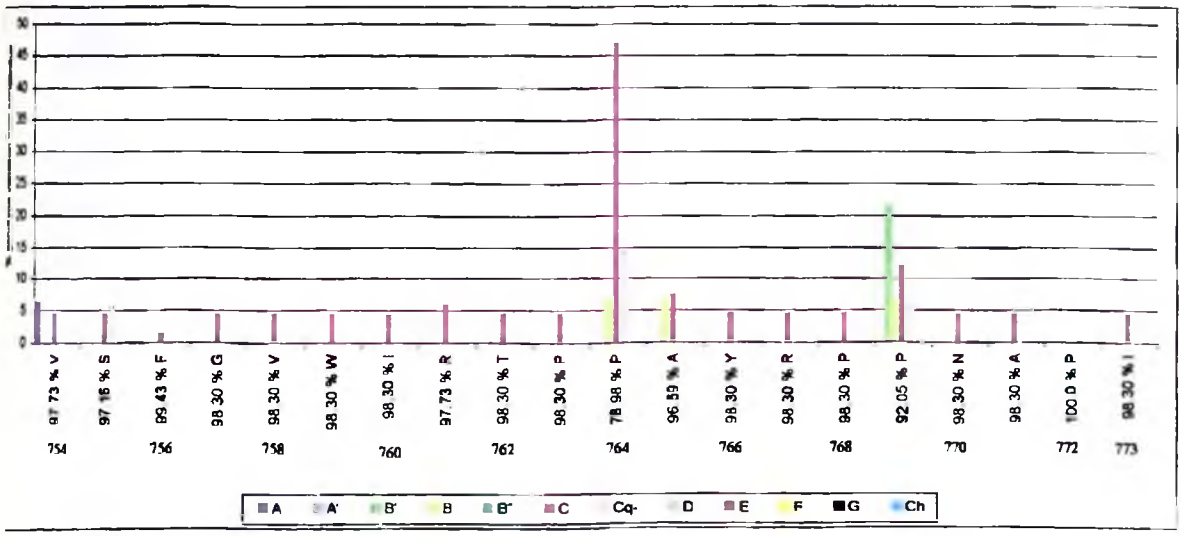


Figure 66. Variation within the published class II DR1-restricted epitope core-120-139

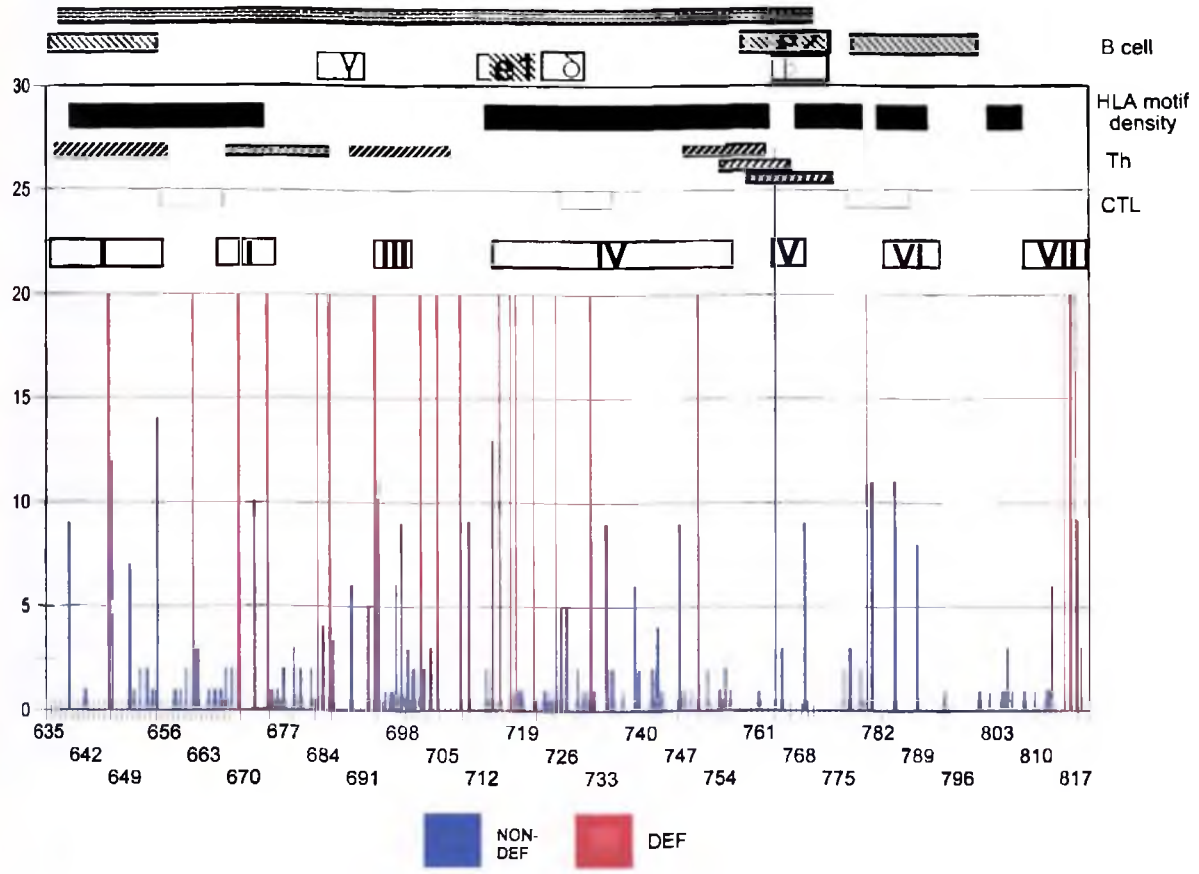


Figure 67. The non-definitive (non-def) and genotypic (def) change within 175 core genes showing defined B and T-cell regions, HLA class I and II motif-dense regions and the 7 heterogenous domains (I-VII) defined by Hosono *et al.* (1995)

at position 764. Of the 64 specimens in the core database with the change at 731, 21 (or 32.8 percent) had the double mutation. Genotype D and E typically did not have the double change. 71.4 percent of the genotype C specimens with either a double or single change were patients with HCC.

The third and last predicted epitope in this cluster extends from RF1 amino acid 743-758 and partially contains three overlapping class II restricted published epitopes: core-111-125 (RF1 745-759, Figure 64), core-117-131 (RF1 751-765, Figure 65) and. core-120-139 (RF1 754-773, Figure 66).

Despite the presence of three published epitopes in this area it is a well conserved stretch of the genome in our database (Figure 67). A non-definitive change at amino acid 747 and the definitive, well conserved change at position 750 are the only changes of note. The change

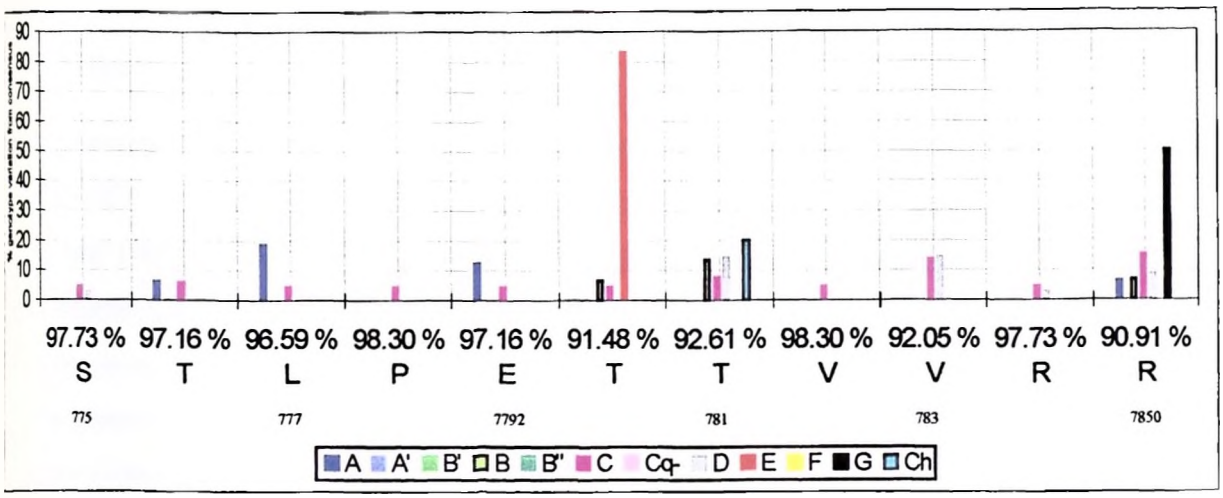


Figure 68. Variation within the published class II DR1 restricted epitope core-141-151

at position 764 which abrogates the secretion abnomality caused by variation at amino acid 731 has already been discussed. It is not surprising that the change at amino acid 764 affects phenotype as it is a change which very weakly conserves the type and properties of the amino acids involved (i.e. proline to glutamine).

10.3.3.5 Potential T-cell epitope region IV (amino acid 765-774)

The third of the overlapping published class II epitopes (754-773) is predominantly within the less dense binding motif region which extends from amino acid 765-774 (Figure 66). In this region amino acid 769, like position 764, has the weakly conserved proline to glutamine or serine change.

10.3.3.6 Potential T-cell epitope region V (amino acid 778-785)

A published class I HLA-A31 and HLA-A68 restricted epitope (Figure 68) extends from RF1 amino acid 775-785 (core-141-151) within the second short region of less dense binding motif concentration towards the 3' end of the core. There are weakly conserved amino acid changes at RF1 position 780 (threonine to asparagine), 781 (threonine to cysteine) and 785 (arginine to cysteine). Only the first of these, the change at amino acid 780 is definitive occuring in genotype E. This genotype E specific variation is outside the genotype D/E recombinant region.

10.3.3.7 Potential T-cell epitope region VI (amino acid 798-802)

No published epitope has been found in the binding motif cluster at the carboxy-terminus of the core gene from RF1 amino acid 798-802.

10.4 Discussion

All genotypes were well conserved over the first 180 nucleotides (approximately nucleotide 1814-1995) varying by only intra-group distances of less than 4 percent (Figure 58). In fact the first 120 nucleotides are so well conserved that the genotypes form a single cluster when phylogenetically compared in this sub-genomic region (not shown). Furthermore this conservation is not only observed within HBV strains but also between hepadna- and retroviruses leading Miller and Robinson (1986) to postulate that the two virus families share a common origin. The last three full intervals of the core (nucleotide 2296-2475) are also well conserved both within the genotypes of HBV and mammalian hepadnaviruses and therefore do not appear to be under considerable selection pressure (Takahashi *et al.*, 2000).

Phylogenetic analysis of 3' core nucleotide data partitions the genotypes into two clusters. The last 21 nucleotides are uniquely conserved within the genotypes A, B, C and D (TCT CGG GAA TCT CAA TGT) on the one hand and genotypes E, F, G and Ch (TCT CCA GCT TCC CAA TGT) on the other. Just upstream of the arginine-rich region at amino acid 191 is the Golgi peptidase motif, TTVVR in all except genotype E (NTVVR). Both GSHV and WHV also have a variation at this position (HTVIR) which may indicate a common progenitor (Takahashi *et al.*, 2000).

All new members of the subgroup B' behaved as recombinants in the core gene region. No new members of subgroup B" were found. It remains to be seen whether the latter two specimens thus represent examples of particularly heterogenous specimens from subgroup B', or whether they are prototypes for a separate subgroup of B. The consistency of the variation in subgroup B' supports our hypothesis that this subgroup of genotype B is an artifact of a fairly common recombination event. Similarly, every newly sequenced member of genotype E had intra-group variation with genotype D at the carboxy-terminal of X, and in the first two-thirds of the precore/core region. Once again the consistency of this finding supports the concept that genotype E has a common recombinant fragment of genotype D, and that southern African genotype E specimens are not different in this respect. This is not a surprising finding as genotype E has only been reported in Africa, and the peoples of southern Africa share a common history with West and East Africans.

Gunther *et al.* (1999) compared wild type sequences from numerous patients including longitudinal studies of single patients and found that about 75 percent of all amino acid changes cluster at 36 "hotspot" positions. Table 24 lists these 36 positions together with an additional 6 sites (amino acid numbers marked with an asterisk in Table 24) where I have found significant variation. The effect of the amino acid changes on the properties of the protein at these 42 sites are variable and are listed in the Table. Changes in the major B cell epitopes are particularly weakly conserved. Forty percent of the changes were definitive changes and 60 percent of the definitive changes would alter the characteristics of the protein. The same proportion of the non-definitive changes were found to affect the properties of the protein.

A previous study has shown that the frequency of core gene mutation is significantly associated with the precore stop codon mutation, hepatitis B e antigen negativity, and active liver disease but is not associated with patient age. Mutations also occur predominantly around the time of HBeAg clearance when liver disease is most active (Akarca and Lok, 1995). After clearance of HBeAg, mutations in regions of the virus recognized by CD4+ helper T-cells and B cells allow persistence of the HBe negative virus and selection of mutations in HBeAb positive patients with viraemia and active hepatitis (Carman *et al.*, 1995).

Figure 67 shows the position of the predicted and published B and T-cell epitopes together with the non-definitive (blue) and definitive (red) change in the core database. Most of the non-conserved change I observed in the core database was within published T helper epitopes (amino acids 639, 647, 655, 685, 697, 764 and 769). There are class I predicted, but not published, epitopes throughout the published class II epitopes. Many of the common variations listed in Table 24 occur at anchor residues (P2 or P9) of these potential epitopes and many of these changes affect the BIMAS score of the epitope. The latter can be calculated using the dissociation tables and constants available at the BIMAS site (Parker *et al.*, 1994 [tables last accessed 27.07.01]). The B cell epitope e1 is weakly conserved at amino acids 708, 713, 715 and 718 and the internal e2 epitope at amino acids 764 and 769. However in the 3' region of the core which contains only published CTL epitopes, the conservation of protein properties is weakly conserved (see RF1 amino acids 780, 781, 785, 789, 791 and 817 in Table 24). Most of this change is also non-definitive change with the

Table 24. Amino acid changes (and their possible effects) in the HBc/e protein sequence

RF1 aa	Consensus	Amino acid conservation	Epitope: Allele specificity	Position within epitope	Effect of change?	Consensus defines	
						groups	sub-groups
639	P→T, H, Q, A	Weak to very weak	CTL(638-646): B51	P2	T, H & Q abrogate epitope. A lowers the BIMAS Score		
			Th(635-654): NS	P5			
			p22-4 Th(636-657): DR17/18		Published class II epitope		
643	T→S, P, A	Moderate	CTL(638-646): B51	P9	S & P lower the BIMAS Score. A has no effect	A, D, & F	B
			CTL(645-653): multiple A & B alleles	P2			
			Th(644-658): DR4	P3			
			p22-5 Th(638-667): DR4				
	S→T, Y	Moderate to weak	CTL(638-646): B51	P9	T & Y raise BIMAS Score	C & F	B & B*
			CTL(645-653): multiple A & B alleles	P2			
			Th(644-658): DR4	P3			
			p22-5 Th(638-667): DR4				
647	V→A, L, T, E, M, G	Weak to very weak	CTL(647-655): A26, B40	P1	A & G raise BIMAS Score, rest have no effect		
			Th(647-661): DR1, DR4	P1			
650	L→ V, I, F	Moderate	CTL(646-650): A24	P9	BIMAS Score lowered	All but F	
			p22-6 Th(648-677)				
			Th(650-664): DR17, DR4	P1			
			Th(647-661): DR1, DR4	P4	Raises the SYFPEITHI Score		
655	S→H, A, V, N, F, T	Very weak to moderate	Th(654-662): A1	P2	Lowers the SYFPEITHI Score		
661	V→I, T	Moderate to weak	CTL(652-661): A2	P10	The published class I epitope has a very high BIMAS Score which is reduced by a factor of 10 when V is switched to either I or T	A, D, E, F	B
			Th(658-672): DR1, DR4, DR11				
			p22-7 Th(656-687): DR4				
665	S→A, T	Moderate	CTL(661-669): B7	P9	A in position P9 results in the highest BIMAS Score, both S & T at this position the lower BIMAS Score	All S but subgroup B has A	
			Th(662-681): DR7	P8	Published Class II epitope		
672	V→F, H, G, N	Y to F strong; Y to H, G or N weak	CTL(666-673): A1, A3 & B62	P8	All changes lower the BIMAS Score markedly		
674	E→D, C, A, K	E to D strong all others weak				All except DEF	Dm
683	S→T, E, Y, F	Moderate to weak	CTL(682-690): A33, A68	P2	S to Y no change, all others lower the BIMAS Score markedly	All except F	
684	F→S, H, N, A	Moderate to weak	Th(684-703): NS	P1	Published Class II epitope		
684	H→N	Weak	CTL(684-692): A*0203	P2	No effect	All but F	
			Th(684-703): published NS	P2			
			P22-9 Th(678-707)				
685	L→I	Strong	CTL(684-689): B7, B60	P9	I lowers BIMAS Score		
693	I→V, G, T	Strong	Th(693-707): DR1, DR4, DR11		No effect	All but Cg-	
			p22-10 Th(688-717): DR3 or DR4				
694	L→V, I	Strong	CTL(686-694): B8, B14, B58, B60	P9	Both V and I lower BIMAS Score		
697	G→N, E, D, V, R	Weak	CTL(696-704): A1	P2	V raises BIMAS Score, N no effect, E, D & R lower BIMAS Score		
			Th(692-711): DR4				
698	E→D	Strong					
701	T→N, S, G	Moderate to weak	CTL(693-707): DR1, DR4	P9	T to N or S no effect. G drops BIMAS Score	A, D, E, F	B
			Th(691-705): DR17	P11	T to N raises the SYFPEITHI Score		

Table 24 continued overleaf

Ref no	Consensus	Amino acid conservation	Epitope: Allele specificity	Position within epitope	Effect of change?	Consensus defines	
						groups	sub-groups
708	V → G, N, A, S,	Weak	e1 B epitope			E	Dm
	S → G, R, N	Weak to very weak	Th(703-717): DR17	P6	Lowers the BIMAS score	C	B1 B1
	N → A, V, T	Weak to very weak				A, F	B
711	E → Q, D, T	Strong to moderate	CTL(710-718): B15, B37, B40, B44, B61	P2	All reduce or abrogate the BIMAS Score	All but subgrp B	Subgrp B = C
			e1 B epitope				
713	P → Q	Weak	CTL(712-720): B7, B8, B14, B51, B52, B61	P2	BIMAS Score is reduced considerably		
			e1 B epitope				
714	A → T, V, S, G, I	Moderate	e1 B epitope				
715	S → A	Weak	CTL(714-722): B35, B44, B58, B62	P2	BIMAS Score is halved in B35, B44 and abrogated in B58	All but F	
			e1 B epitope				
717	D → E	Strong	Th(717-731): DR1	P1	No effect		
			e1 B epitope				
718	L → R, V, A, G, D	Weak to very weak	CTL(710-718): B15, B37, B40, B44, B60, B61	P9	BIMAS Score lowered by V, A and abrogated by Q, R & D		
721	S → N, G, T	Moderate to weak	CTL(721-734): A11		Published epitope	C, D, E	B, B1
			e1 B epitope				
725	T → V, N, A	Moderate to weak	CTL(724-732): A31, A*6801	P2	V raises the BIMAS Score but N & I lower the BIMAS Score	A, D, E, F	B
			CTL(717-725): A26	P9	No effect		
			Th(717-731): DR4	P9	Lowers the SYFPEITHI Score		
726	N → T, H	Weak	Th(726-740): DR3				
			p22-13 Th(718-747)	P9			
727	N → V, T, L	Moderate	CTL(719-727): B7	P9	L raises BIMAS Score but V & T abrogate response		
			Th(720-738): DR17				
731	I → L, F	Moderate to weak	CTL(723-731): B51	P9	I to L lowers BS but F abrogates the response	A, B, C, F	
	F → I	Weak	CTL(723-731): B51		I reconstitutes a response	D, E	
734	L → I	Strong	CTL(733-741): A2, A3, B63	P2	L to I lowers binding in A2 & B62 but abrogates the A3 response		
			CTL(726-734): A2, B8, B15	P9	BS are not high with L but I lowers them further		
			P22-14 Th(728-757): DR4, DR11				
			Th(733-747): DR11				
735	I → T, V, L	Moderate to weak	CTL(731-739): B27	P9	L no effect, T & V lower BS		
			CTL(738-746): A3, A11, A31, A33, A*6801	P2	All variants respond but optimum is variable		
			Th(739-753): DR3				
747	E → Q, C, S	Moderate to weak	Th(745-758): DR1	P3			
			Th(747-761): DR4, DR1	P1			
			CTL(746-754): B37, B40, B44, B52, B60, B61	P2	All variants have a positive BS but optimum amino acid varies with different alleles		
750	I → L, V	Strong conservation	Th(745-758): DR1	P6	Published epitope	D, E	
			CTL(742-750): A2, A26, B8, B14, B15, B38B39, B52, B58	P9	Change to V or I lowers BS in all except B58 where it raises BS		
			p22-16 Th(748-777)				
			Th(755-769): DR3				
764	D → Q, I, L, A, N, Sc	Weak to very weak	CTL(763-761): B7	P2	All change lowers BS	All but E	
			Th(751-765): DR6		Published epitope		
			e2 B epitope				
			p22-17 Th(758-789) & p22—18 (768-792)				
			Th(771-785): DR3, DR4, DR11				

Table 24 continued overleaf

RF1 aa	Consensus	Amino acid conservation	Epitope: Allele specificity	Position within epitope	Effect of change?	Consensus defines	
						groups	sub-groups
766	P →A, Q, L, S, T	Weak to very weak	CTL(768-776)	P2	L raises BS all others decrease BS		
			e2 B epitope				
			Th (754-773): DR3				
780	T →N	Weak	CTL(775-785): A31, A68		No effect	All but E	
781	T →C, S, A	Moderate to very weak	CTL(780-788): A3A68	P2	For A68 T has highest BS but for A33 S has the highest BS		
785	R →C, G, C, P, E	Very weak	CTL(777-785): A1, A33	P9	All variants abrogate the response		
4785	d →C, D, M, R	Very weak					
4791	S →T, W, P	Moderate to very weak	CTL(790-798): B14	P2	P & T no effect, W abrogates responseP & T no effect, W abrogates response		
			p22-20 Th(790-819)				
4817	S →P	Very weak					

* marks sites found to be variable in our analysis and not reported by Gunther *et al.*, 1999
# Numbers prefaced with a hatch correspond to numbers of Gunther *et al.*, 1999 less two since, for consistency, we have taken into consideration the two bp insertion at 787-788 in genotype A only.
P1-P9 etc. indicates the position of the variant site in published and/or predicted epitopes
NS denotes published epitopes for which the allele specificity is not stated

{ The large brackets mark the possible epitopes to which the patients' responded

exception of the change at positon 780 which is a genotype E specific change which could affect the immune response of southern African specimens belonging to this genotype. Most of the changes in B' and B" versus subgroup B itself would result in immune escape. Two of the amino acids in the e2 B epitope also differentiate the B' and B" subgroups from subgroup B. Genotype F also had several unique changes most of which would allow immune escape.

Several local specimens sequenced had premature stop codons although these were not the common precore stop codon (ES05 RF1 amino acid 633, JB11 RF1 amino acid 653, ES04 and ES05 RF1 amino acid 680). Mutant strains are escaping from the immune response rather than inducing a fresh immune response since core mutations do not lead to viral clearance, and most substitutions occur in the initial burst of hepatic activity and are not transient (Carman *et al.*, 1997). Mutations in the T helper epitopes allow immune escape and minimise immune-mediated hepatitis manifested by remission in the patient. Thus as demonstrated by Dusheiko *et al.* (1985), our results also indicate that the southern African patient has predominantly asymptomatic disease.

The core gene has both class I and class II binding motif-rich regions. Epitopes were predicted between RF1 amino acids 638-672, 684-705, 709-761, 765-774 and 778-785

(Figures 52 and 54 in Chapter 9) and published epitopes were largely found within these motif dense regions. Hosono *et al.* (1995) defined three major (I, IV and V) and four minor (II, III, VI and VII) domains of major non-definitive change in a group of tumour patients. The position of these domains has been included in Figure 67 and several of these domains coincide with T-cell regions, and/or motif-rich binding regions, in the core protein.

The conservative change observed may also have some effect on the immune response. Bertoletti *et al.* (1994) investigated the effect of variant peptides which decreased either the binding affinity or the TCR recognition, or both, and found that some clones had strong antagonistic effects at all concentrations while others were pure agonists. Still others showed antagonism at low concentrations with decreasing inhibition at higher concentration with the appearance of an agonist effect.

CHAPTER 11

FINAL DISCUSSION AND FUTURE DIRECTION

HBV is hyper-endemic in most parts of sub-Saharan Africa. Since there is a growing perception that genotypic differences in HBV are related to chronic pathogenesis in the infected liver and little is known of the HBV strains circulating in southern Africa, I attempted to better characterize both HBV and the African host in this part of the world. Although five of the 7 genotypic groups (A, B, C, D, E) were found in urban Gauteng, invariably genotype B and C specimens were only found in Africans of Asian descent and were therefore probably imported. No genotype F or G has been found locally to date. Only genotypes A (69.6 percent in urban Gauteng and 89.2 percent in a rural area in the Eastern Cape) and D (16.1 percent in Gauteng and 10.8 percent in the Eastern Cape) are found in significant numbers in South Africa. In Namibia, genotype E was the predominant genotype and was found in 93.8 percent of specimens examined together with 6.2 percent of genotype D.

Genotypic screening revealed that 64.1 percent of urban and 97.0 percent of rural African genotype A specimens did not have the preS1 and preS2 sites predicted from GenBank genotype A sequence data. Sequencing and phylogenetic analysis revealed that the aberrant specimens differed by more than 4 percent from the well conserved genotype A specimens found in other parts of the world. They clustered together and away from global genotype A specimens to form a subgroup which I have called subgroup A' (Bowyer *et al.*, 1997). The amino acid differences in the preS2/S region that set these isolates apart from the rest of genotype A were at preS2 amino acids 7, 10, 32, 35, 47, 48, 53, and 54 and S amino acid sites 207 and 209. Sequence data emerging recently from Central and South America shows that this subgroup of genotype A also occurs in the New World. Local genotype E specimens do not differ significantly from published genotype E sequence data. This was not unexpected as this genotype originates from, and is unique to, Africa.

Two specimens from the new genotype of HBV, genotype G, which was first identified last year (Stuyver *et al.*, 2000) were found to cluster together and away from all other genotypes

in the phylogenetic trees of our 1997 preS2/S study. I listed the characteristic preS2/S amino acids which set them apart from the rest of the HBV specimens (Figure 23; Bowyer *et al.*, 1997) although I did not go so far as to suggest that they were representatives of a new genotype.

Both genotypic and serotypic analyses of southern African isolates showed that there was a much higher prevalence of the *ay* determinant in South Africa than previously reported ranging from 20 percent in the preS2/S sequence study to more than 44 percent in Hardie's study from the Western Cape. The sera used in the previous serotypic study (Courouce-Pauty *et al.*, 1983) came from black donors attending the Natal Blood Transfusion Service in the Eastern Kwazulu-Natal province of South Africa, and may reflect a regional difference in genotype distribution which still needs to be investigated.

Since the genotypes of HBV are geographically distributed variation should be globally distributed as well. By defining the "hotspots" within the different genotypes and differentiating between "definitive or genotypic" change and "non-definitive or random" change within each genotype, I further characterized this variation. This data serves as baseline data for other studies. For example, although amino-acid 19 of RF1 is one of four amino acids thought to be responsible for the universal "*b*" determinant used to identify specimens with sufficient preS2 protein to enable them to be typed by HBV genotype EIA (Usuda *et al.*,1999), it was found to be mutated in 17 percent of all specimens examined and 56 percent of the genotype A specimens.

I also compared the total number of random changes within the various domains of HBV and sorted specimens according to amount of non-definitive change. The core gene was the only domain where random change appeared to be directly related to the stage and degree of liver disease (Figure 39). The number of amino acids changes was also related to a higher incidence of the 1896 premature precore stop mutation. This study also identified subtype variation at anchor residues in 12/41 published epitopes (see Table 2 for a list of published epitopes). Some of these such as $^{(RF1\ 43)}$Phe → Ser, $^{(RF1\ 73)}$Leu → Ser, $^{(RF1\ 94)}$Leu → Ile/Pro/Arg, $^{(RF1\ 101)}$Leu → Pro, $^{(RF1\ 245)}$Ser → Leu, $^{(RF2\ 565)}$Asp → Try, $^{(RF2\ 840)}$Val → Cys/Ser, $^{(RF3\ 990)}$Asn → Ala/Arg and $^{(RF3\ 1042)}$Ala → Ile/Asn/Asp alter protein properties and could allow immune escape and thereby minimize immune-mediated hepatitis. While their role as

antogonists has not been fully explored, other conservative changes such as $^{(RF1\ 50)}$Ala → Val, $^{(RF1\ 261)}$Leu → Val, $^{(RF1\ 661)}$Val → Ile, $^{(RF1\ 750)}$Leu → Ile, $^{(RF2\ 574)}$Val → Leu and $^{(RF3\ 1035)}$Ile → Met/Val would not be expected to drastically affect binding and therefore the immune response.

I noticed species which clustered in different genotypes when different parts of the HBV genome were used for phylogenetic analyses, and that the variation within these specimens was mosaic in nature. This motivated the search for evidence of recombination in HBV (Chapter 7). Our investigation showed that putative A/D, A/E and B/C recombination was fairly common in HBV infection (Bowyer *et al.*, 2000), and recently several groups have corroborated our results and report evidence of recombination within HBV isolates (Yuasa *et al.*, 2000; Hannoun *et al.*, 2000; Owiredu *et al.*, 2001). Since recombination implies infection of the individual by more than one genotype this subject, once thought to be unimportant, needs to be explored. Furthermore, care must be taken in studies, such as this one, to include full-genome sequence results, or at least sequence data from more than one region of the genome, in order to be sure that one is not dealing with a recombinant. Two earlier horizontal studies suggested a genotype switch from genotype A to genotype D during disease progression (Gerner *et al.*, 1998; Bahn *et al.*, 1997), but these authors may also be observing the results of recombination since neither of these studies sequenced the full genome.

The histograms showing the distribution of average pair-wise distances (Figure 45) between the genotypes over subgenomic regions of HBV indicate that substitution rates are not constant among the lineages in the preS regions and between genotype F and all other genotypes, A to E, over most of the genome. This is contrary to the findings of Yang *et al.* (1995), but genotypes E and F and subgroups B, B' and Cq' were not represented in their study.

Although the ends of some of the mosaic fragments correspond approximately to the end of the major epitopes of the core gene, many epitopes of P and all preS1 epitopes are not represented in the putative recombinant fragments. I therefore conclude that different MHC backgrounds in human populations in different parts of the world are unlikely to be entirely responsible for the apparent recombination of HBV fragments which I have observed.

Further study is required to produce direct evidence for recombination in HBV, and to clarify the role of recombination in the heterogeneity of HBV since the mosaic structure which I and others have observed affects entire clades and alters the phylogeny of HBV over extensive subgenomic fragments.

When the number of genotype B specimens included in the phylogenetic trees was increased from 12 in the original study (Chapter 7) to 31 in the core gene variation study (Chapter 10), subgroup B' expanded from 4/12 (33.3 percent) to 14/31 (45.2 percent) of all genotype B specimens and all new members of the subgroup B' behaved as recombinants in the core gene region. No new members of subgroup B" were found. It remains to be seen whether the latter two specimens thus represent examples of particularly heterogenous specimens from subgroup B' or whether they are prototypes for a separate subgroup of B. The consistency of the variation in subgroup B' supports our hypothesis that this subgroup of genotype B is an artifact of a fairly common recombination event. Similarly every newly sequenced member of genotype E had intra-group variation with genotype D in the carboxy-terminal of X and in the first two-thirds of the precore/core region. Once again the consistency of this finding supports the concept that genotype E has a common recombinant fragment of genotype D sequences and that southern African genotype E specimens are not different in this respect. This is not a surprising finding as genotype E has only been reported in Africa, and the peoples of southern Africa share a common history with West and East Africans.

Population frequency data shows that the five most common HLA-DR genetic types are the same in populations representative of the Western (Zulu) and Eastern (Sotho) Bantu migrations and include two of the alleles, DR2 and DRB1*13, which have been found to be protective against HBV infection. The former protects against both transient and persistent infection while the latter was found to protect against persistent infection. Diepolder *et al.* (1998) suggest that the beneficial effect of the HLA-DR13 allele on the outcome of HBV infection is due to more proficient antigen presentation by the HLA-DR13 molecules leading to a more vigorous HBV core-specific CD4+ T-cell response. There is evidence for the selection of the protective DRB1*13 allele in the West African host (Thurz *et al.*, 1995). In agreement with published data relating host type to disease progression, only one allele of one patient among the 11 patients which I typed had either of the protective alleles. Less

than 48 percent of Bushmen were found to have evidence of past HBV infection (Aspinall *et al.*, 1994) which is much less than the South African and the Namibian average of HBV viral markers (76 and 87.5 percent, respectively), but contrary to what would be expected the DR2 allele is only present in 1.9 percent of the San population as opposed to 11.4 percent of the Shona and Zulu populations. The frequency of the DR13 allele in the San (24.5 percent) is approximately 1.5 times greater than that found within their Bantu neighbours. Bushmen with this allele would be protected against HBV persistence and the rate of chronic carriage is reported to be lower in the San (7.3 percent) than in South Africans (9.6 percent).

CAH is not a common disease manifestation in southern Africa and most South African adult HBV carriers progress at a young age to silent predominantly non-replicative disease. Accordingly, HLA–B35 which has been associated with CAH is a rare allele in southern and central Africa whilst one of the definitive African alleles, B15, has been associated with asymptomatic HBsAg carriage. B15 is present in 17.3 percent of Xhosa and 24.5 percent of the Zulu. Furthermore, the dominant African DQ allele, with a frequency of 40 to 50 percent in all the Bantu populations, is the allele DQ1 which is reportedly found less frequently in CAH (van Hattum *et al.*, 1987). The vaccinated individual and all of the patients with acute hepatitis B had at least one DR3 allele (ES06, ES15, ES16, HB05, SB). This allele is also associated with more active disease. However, three of the chronic carriers (N231, N232 and N238) also had the DR3 allele although it is perhaps significant that two of the three chronic carriers with the DR3 allele appeared to be homozygous for DR. Heterozygosity at the DR and DQ class II loci has been found to be associated with protection from persistent infection (Thursz, 1997b) while homozygosity for HLA class I (Pollicino *et al.*, 1996) and class II (Thio *et al.*, 1999) has been found to be responsible for an insufficient response to HBV, favouring chronic outcome of the infection and long-lasting persistence. Only one allele was detected in 4/7 chronic carriers tested. Clearly more southern African data is required as the entire haplotype of the host is needed to understand the effect of HLA type on disease outcome. In addition, as molecular typing methods supercede old serological typing methods, high resolution population studies need to be performed both here and elsewhere since specific alleles of the same HLA antigen can predispose towards different disease outcomes. This distinction cannot be observed using low resolution serological data.

Proliferation responses to overlapping core peptides were fairly weak in the southern

African patients tested but only four of these had acute disease. The rest of the patients were asymptomatic nonviraemic HBsAg-positive carriers with normal aminotransferase levels and I therefore expected them to have weak proliferation responses. Ninety percent of South African HBV chronic carriers fit this profile. All of the acute patients responded to more than one region of the core while only one chronic carrier, N232, had a response to more than one peptide, and only three of the chronic carriers responded at all. This is in agreement with the finding that chronic carriers have weak and more narrowly focussed T-cell responses while patients who clear the disease have more vigorous, multispecific responses. The immunogenic regions of the genome in southern African patients were found to be essentially the same as published data (Figure 54), but the epitopes eliciting the responses were specific for the African host. The published core-1-20 DR specific epitope appeared to be DR17 and DR18 restricted. Proliferation responses indicated that there were two unpublished DR4-restricted epitopes, one within our p22-5 peptide (RF1 amino acid 638-667) and another within the p22-6 peptide (RF1 amino acid 658-687) and at least one DR17-restricted epitope within p22-13 (RF1 amino acid 718-747). Three overlapping published DR1 or DR6-restricted epitopes are reported in the region of the core which I found to be the most immunogenic. However, none of these alleles were found in any of our patients. All responders to this part of the genome had the broad DR3 allele, but three other DR3 typed specimens did not respond. Two of the 3 non-responders shared the DR18 split of DR3. This is an example of an immunogenic region with different specificity in different populations.

The finding of different immunodominant target molecules in different populations is also to be expected as the frequency of the various HLA alleles is ethnically distributed. For example, published data reports class II linear epitopes in the core region which are all either DR1, DR6 or DR7 restricted and these alleles are present in the Caucasian population at a 10.7, 13.0 and 20.8 percent frequency, respectively; this compares to frequencies of 2.9, 20.7 and 11.5 percent, respectively, in the Sotho and 5.5, 13.3 and 0.4 in the Japanese. Two of our specimens both with the broad DR3 specificity responded to both p22-17 and p22-18 which could mean that the immunogenic region was within the overlap of these two peptides i.e. between RF1 amino acid 768 and 789. One specimen reacted to p22-20 a peptide spanning RF1 amino acid 790-820 at the carboxy-terminal of the core a region which has not previously been reported to be immunogenic.

Many published class I-restricted epitopes are specific for the HLA-A2 allele including

core-18-27 (Table 2). The HLA-A*0201 allele is the most common HLA-A2 allele in many populations but notably not among the South African Zulus (Figure 55), in whom the most frequent A2 allele is A*0202. Allele A*0202 was found to have only eleven MHC-binding motifs in the core gene and these all had very low binding scores whereas the more common allele, A*0201, has a total of 202 binding motifs most of which had very high binding scores. The Eastern part of South Africa which includes Kwazulu-Natal, home of the Zulus, has the highest carrier rate in South Africa (Figure 16).

In general, I could correlate DR type with predicted or published epitopes in regions eliciting an immune response in our study and the responses observed were, for the most part, in regions of high motif density. Published motifs also clustered within the regions of high motif density with the exception of three of the DR specific epitopes which overlapped the predicted epitope regions. This is probably a result of the relative lack of DR-favoured motifs in the SYFPEITHI database.

Gunther *et al.* (1999) compared wild type sequences from numerous patients including longitudinal studies of single patients and found that about 75 percent of all core gene amino acid changes cluster at 36 "hotspot" positions. I found significant variation at these 36 positions and an additional 6 sites as well. Most of the non-conservative change was in the Th and B cell epitopes with the exception of the 3' part of the core where non-conservative change was at CTL epitopes. Changes in the major B cell epitopes is particularly weakly conserved. This is in agreement with the finding of Carman *et al.* (1995; 1997). Forty percent of the changes were at definitive sites which would make them genotype specific changes and approximately 60 percent of the definitive changes would alter the characteristics of the protein. The same proportion of the non-definitive changes would be expected to affect the properties of the protein. Mutations that abrogate recognition of a viral epitope by class I-restricted CTL are unlikely to lead to viral escape in patients where the CTL response is simultaneously directed against multiple viral epitopes, as has been reported for patients with acute self-limited HBV infection, but might contribute to viral persistence in a subset of patients with chronic HBV infection who express a narrow repertoire of anti-HBV CTL responses (Bertoletti *et al.*, 1994).

The precore premature stop mutation is one of the variations which has been found to be

genotype dependent. The mutation is rare in genotype A specimens while 50 percent, or more, of subtype D specimens have been found to have this change. It was present in only two of the southern African specimens sequenced in the core gene region, one of these was from genotype C and the other from genotype E. Three other local specimens, two from genotype E and one from subgroup A', had premature stop codons at one or other position of the core which would result in a truncated core antigen. Various groups have speculated (Valliammai *et al.*, 1995; Kramvis *et al.*, 1997) that the G to T mutation at nucleotide 1862 could also be responsible for a disruption in HBeAg expression leading to an HBeAg negative phenotype. This mutation is also genotype dependent as it is definitive of subgroup A'.

While many nucleotide changes have never or rarely been found in HBV isolates from patients without fulminant hepatitis, a distinct mutation present in all genomes has also not been identified (Sterneck *et al.*, 1996), and the same virus can cause asymptomatic disease in one patient while manifesting as fulminant hepatitis in another (Horikita *et al.* 1994). Host genetic type is integral to our understanding of HBV infection and patient response to therapy. However, recently several studies have reported evidence which associates genotype with clinical response. Kao *et al.* (2000) and Ding *et al.* (2001) found that genotype C is associated with more severe liver disease than genotype B in older Taiwanese and Chinese patients, respectively. The former study also found, however, that genotype B was associated with the development of HCC in young Taiwanese. Orito *et al.* (2001) found longer duration of infection in association with genotype C in Japanese patients and that this was associated with a predisposition in these patients to the double mutation in the core promoter (T^{1762}/A^{1764}). Therapeutic responses have also often been related back to the occurrence of specific mutations, such as the T^{1762}/A^{1764} core promoter mutation or the precore stop mutation. Both of these mutations have been found to affect the patient response to interferon (Fattovich *et al.*, 1995; Erhardt *et al.*, 2000; Hunt *et al.*, 2000). In most patients lamivudine treatment induces marked reduction of serum HBV DNA levels and improves serum aminotransferase activities as well as hepatic histology but continued therapy has been found to lead to viral resistance in up to 20 percent of patients within a year, and even more with longer term therapy (Kao *et al.*, 2002). Recently Zollner *et al.* (2001) reported that lamivudine resistance is subtype dependent and that HBV serotype *adw* (exclusive to genotype A in Europe and Africa) is associated with a higher risk of

lamivudine resistance than serotype *ayw* (the serotype of all genotype D and E specimens). Thus it is important that studies continue to document the genetic type of the host and local HBV viral strains, particularly in areas such as sub-Saharan Africa where the virus is endemic. Automated sequencers enable large scale sequence analysis for genotype determination and these methods have the added advantage of enabling constant surveillance for the emergence of mutants under the selection pressures of universal HBV vaccination. With the success of the vaccine, strategies, including the development of therapeutic vaccines, for successful treatment of the millions of chronic carriers worldwide are becoming a priority in HBV research.

In the last few years wild type chimpanzee and other Old and New World primate specific hepatitis B viruses have been identified. These include wild type hepadnaviruses isolated from the the gibbon (Norder *et al.*, 1996; Lanford *et al.*, 2000), the woolly monkey (Lanford *et al.*, 1998), the orangutan (Warren *et al.*, 1999) and the chimpanzee (Hu *et al.*, 2000; Takahashi *et al.*, 2000; MacDonald *et al.*, 2000). Although the first chimpanzee virus was reported by Vaudin *et al.* in 1988, the animal was a resident of the London Zoo and the authors were not sure whether the virus represented an adaptation of the human virus to a chimpanzee host or the first reported complete sequence of an African human or non-human primate virus. The recent finding of similar viruses in wild caught non-human primates, however, confirmed that the virus was indeed chimpanzee-specific and this has again sparked interest in the origins of HBV. Bollyky *et al.* (1999) proposed that HBV originated in the New World and spread from the Americas into Europe and elsewhere in the Old World after contact between Europeans and indigenous peoples up to 400 years ago. Since the level of sequence divergence between the primate specific viruses is similar to that found among human HBV genotypes A to G, this theory of the origin of HBV seems extremely unlikely.

It has also been proposed that HBV infection was present in anatomically modern humans 100-150,000 years ago and has evolved geographically since the global dispersal of humans. This too is unlikely since, with the exception of the variant found in woolly monkeys, the various nonhuman primate viruses are interspersed among the human genotypes in the phylogenetic tree (MacDonald *et al.*, 2000).

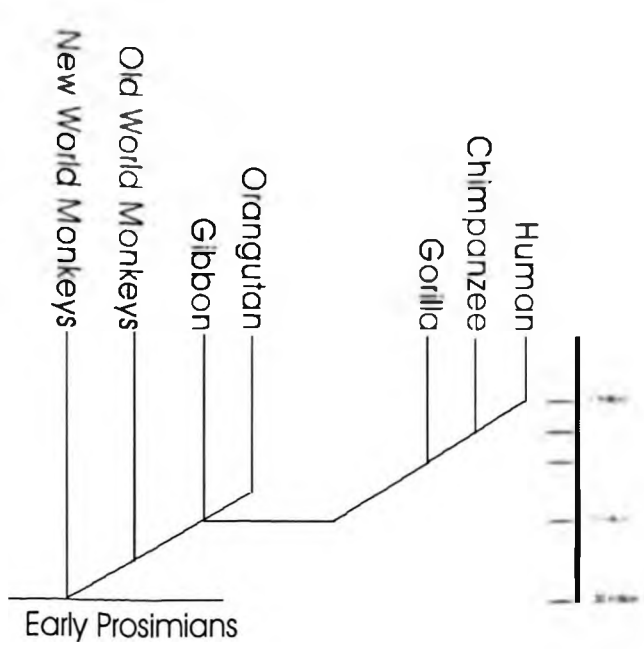


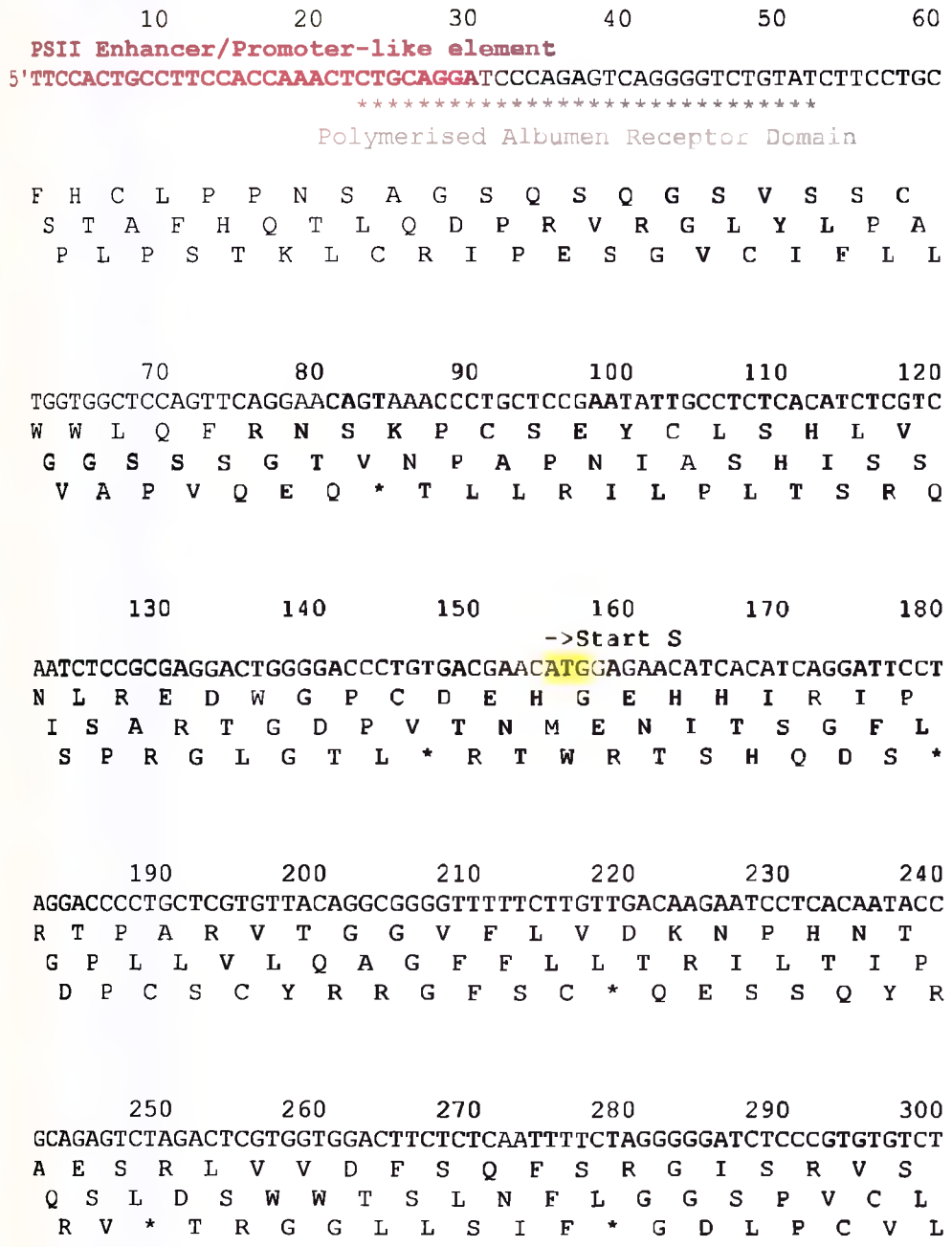
Figure 69. Evolution of modern primates (adapted from Microsoft Encarta, 1995)

The phylogeny of the host species would reflect the phylogenetic tree of the various primate HBV variants if the HBV viruses have co-evolved in their primate hosts over periods of 10 to 35 million years with the numerous genotypes found in humans originating through multiple zoonotic transmission episodes from several nonhuman primate species infected with different species-specific genotypes as described for HIV (MacDonald *et al.*, 2000). The phylogenetic tree of the various primate HBV variants does not, however, reflect the phylogeny of the host species (Figure 69). For example, the gibbon and orangutan of Asia are widely separated from the chimpanzee of Africa but published trees show the gibbon and chimpanzee HBV nucleic acid sequences more closely related to each other than to the human genotypes. Only the woolly monkey HBV and human genotype F appear to be related with respect to their phylogenetic and geographic (South American) proximity (Hu *et al.*, 2000).

The discovery of new primate species or new HBV genotypes would of course alter the phylogeny of their respective trees. To clearly establish the link between chimpanzee and human HBV and to clarify the evolutionary history of HBV in human populations, areas of high endemicity such as southern Africa need to be thoroughly sampled to rule out the existence of other genotypes, particularly since disease outcome has been associated with genotype. To this end, it is important that molecular epidemiological data be gathered from the many parts of Africa from which no data exists, including South Africa's eastern neighbour, Mozambique, where both HBV and HCC are endemic and widespread.

APPENDIX A

THE POSITION OF ORFs AND BINDING SITES ON THE HBV GENOME SHOWN ON GENBANK SEQUENCE ACCESSION NUMBER X02763



```
      310       320       330       340       350       360
                                    Binds GRE and operates with ENHI
                                    GRE consensus: NCAANNTGT
TGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTCACCAACCTCCTGTCCTCCAATTTG
W  P  K  F  A  V  P  N  L  Q  S  L  T  N  L  L  S  S  N  L
 G  Q  N  S  Q  S  P  T  S  N  H  S  P  T  S  C  P  P  I  C
  A  K  I  R  S  P  Q  P  P  I  T  H  Q  P  P  V  L  Q  F  V


      370       380       390       400       410       420
TCCTGGTTATCGCTGGATGTGTCTGCGGCGTTTTATCATATTCCTCTTCATCCTGCTGCT
S  W  L  S  L  D  V  S  A  A  F  Y  H  I  P  L  H  P  A  A
 P  G  Y  R  W  M  C  L  R  R  F  I  I  F  L  F  I  L  L  L
  L  V  I  A  G  C  V  C  G  V  L  S  Y  S  S  S  S  C  C  Y


      430       440       450       460       470       480
ATGCCTCATCTTCTTATTGGTTCTTCTGGATTATCAAGGTATGTTGCCCGTTTGTCCTCT
M  P  H  L  L  I  G  S  S  G  L  S  R  Y  V  A  R  L  S  S
 C  L  I  F  L  L  V  L  L  D  Y  Q  G  M  L  P  V  C  P  L
  A  S  S  S  Y  W  F  F  W  I  I  K  V  C  C  P  F  V  L  *


      490       500       510       520       530       540
                                    d/y
AATTCCAGGATCAACAACAACCAGTACGGGACCATGCAAAACCTGCACGACTCCTGCTCA
N  S  R  I  N  N  N  Q  Y  G  T  M  Q  N  L  H  D  S  C  S
 I  P  G  S  T  T  T  S  T  G  P  C  K  T  C  T  T  P  A  Q
  F  Q  D  Q  Q  Q  P  V  R  D  H  A  K  P  A  R  L  L  L  K


      550       560       570       580       590       600
          <-aa 139-147 is often referred to as the "a" epitope->
AGGCAACTCTATGTTTCCCTCATGTTGCTGTACAAAACCTACGGATGGAAATTGCACCTG
R  Q  L  Y  V  S  L  M  L  L  Y  K  T  Y  G  W  K  L  H  L
 G  N  S  M  F  P  S  C  C  C  T  K  P  T  D  G  N  C  T  C
  A  T  L  C  F  P  H  V  A  V  Q  N  L  R  M  E  I  A  P  V


      610       620       630       640       650       660
                                 r/w
TATTCCCATCCCATCGTCCTGGGCTTTCGCAAAATACCTATGGGAGTGGGCCTCAGTCCG
Y  S  H  P  I  V  L  G  F  R  K  I  P  M  G  V  G  L  S  P
 I  P  I  P  S  S  W  A  F  A  K  Y  L  W  E  W  A  S  V  R
  F  P  S  H  R  P  G  L  S  Q  N  T  Y  G  S  G  P  Q  S  V
```

```
      670        680        690        700        710        720
TTTCTCTTGGCTCAGTTTACTAGTGCCATTTGTTCAGTGGTTCGTAGGGCTTTCCCCCAC
F  L  L  A  Q  F  T  S  A  I  C  S  V  V  R  R  A  F  P  H
 F  S  W  L  S  L  L  V  P  F  V  Q  W  F  V  G  L  S  P  T
  S  L  G  S  V  Y  *  C  H  L  F  S  G  S  *  G  F  P  P  L

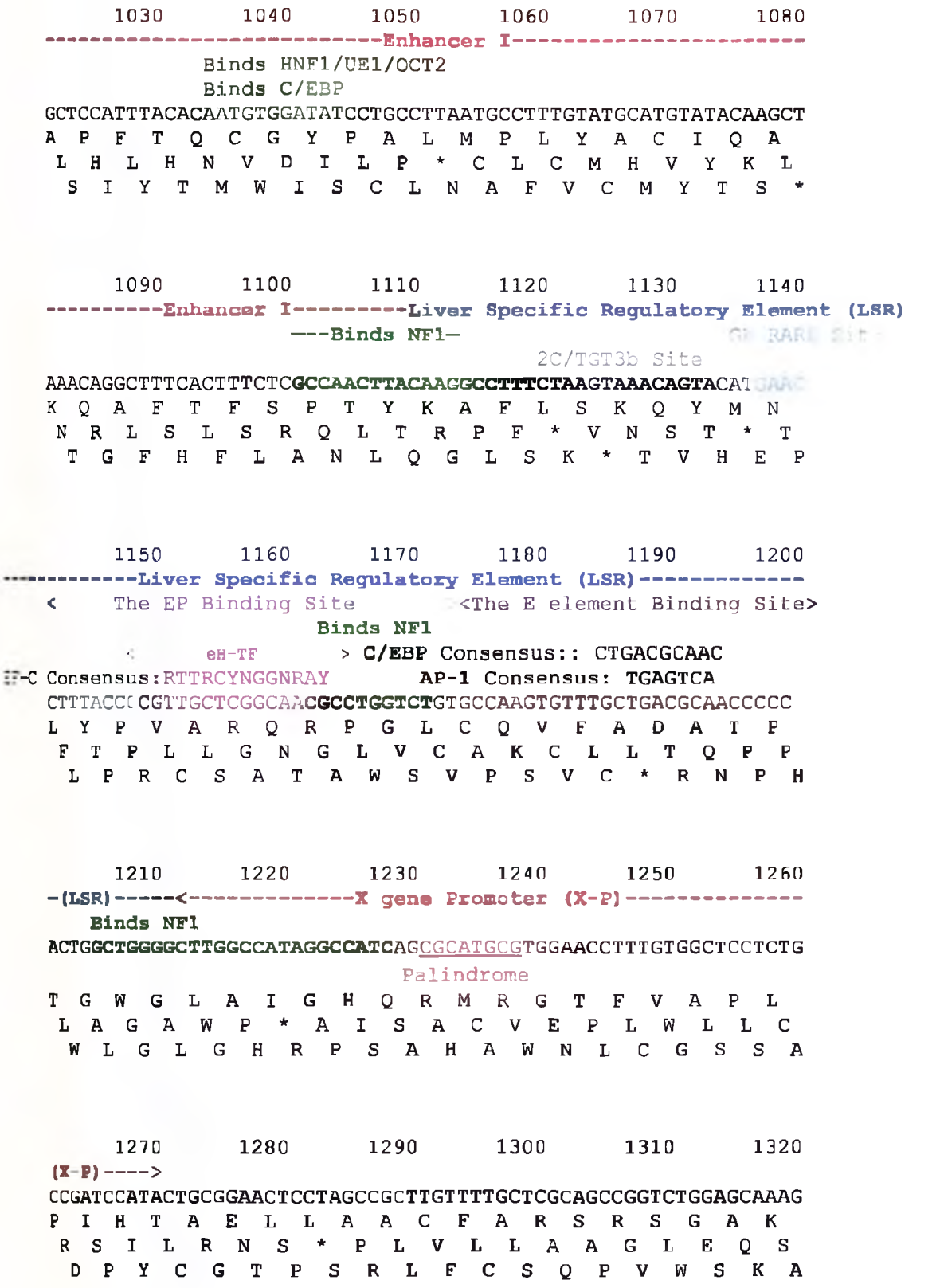
      730        740        750        760        770        780
TGTTTGGCTTTCAGCTATATGGATGATGTGGTATTGGGGGCCAAGTCTGTACAGCATCGT
C  L  A  F  S  Y  M  D  D  V  V  L  G  A  K  S  V  Q  H  R
 V  W  L  S  A  I  W  M  M  W  Y  W  G  P  S  L  Y  S  I  V
  F  G  F  Q  L  Y  G  *  C  G  I  G  G  Q  V  C  T  A  S  *

      790        800        810        820        830        840
                                                   End S
GAGTCCCTTTATACCGCTGTTACCAATTTTCTTTTGTCTCTGGGTATACATT[redacted]ACCCT
E  S  L  Y  T  A  V  T  N  F  L  L  S  L  G  I  H  L  N  P
 S  P  F  I  P  L  L  P  I  F  F  C  L  W  V  Y  I  *  T  L
  V  P  L  Y  R  C  Y  Q  F  S  F  V  S  G  Y  T  F  K  P  *

                                                <-Enhancer I-------
      850        860        870        880        890        900
                                                -----Binds UE3-----
AACAAAACAAAAAGATGGGGTTATTCCCTAAACTTCATGGGCTACATAATTGGAAGTTGG
N  K  T  K  R  W  G  Y  S  L  N  F  M  G  Y  I  I  G  S  W
 T  K  Q  K  D  G  V  I  P  *  T  S  W  A  T  *  L  E  V  G
  O  N  K  K  M  G  L  F  P  K  L  H  G  L  H  N  W  K  L  G

      910        920        930        940        950        960
-------------------------Enhancer I----------------------
GGAACTTTGCCACAGGATCATATTGTACAAAAGATCAAACACTGTTTTAGAAAACTTCCT
G  T  L  P  Q  D  H  I  V  Q  K  I  K  H  C  F  R  K  L  P
 E  L  C  H  R  I  I  L  Y  K  R  S  N  T  V  L  E  N  F  L
  N  F  A  T  G  S  Y  C  T  K  D  Q  T  L  F  *  K  T  S  C

      970        980        990       1000       1010       1020
-------------------------Enhancer I----------------------
                Binds UE3
                     Binds C/EBP
  Consensus of C/EBP:    RTTGCGYAAY
GTAACAGGCCTATTGATTGGAAAGTATGTCAAAGAAITGTG GTCTTTTGGGCTTTGCT
V  N  R  P  I  D  W  K  V  C  Q  R  I  V  G  L  L  G  F  A
 L  T  G  L  L  I  G  K  Y  V  K  E  L  W  V  F  W  A  L  L
  *  Q  A  Y  *  L  E  S  M  S  K  N  C  G  S  F  G  L  C  C
```


1030 1040 1050 1060 1070 1080
Enhancer I
Binds HNF1/UE1/OCT2
Binds C/EBP
GCTCCATTTACACAATGTGGATATCCTGCCTTAATGCCTTTGTATGCATGTATACAAGCT
A P F T Q C G Y P A L M P L Y A C I Q A
L H L H N V D I L P * C L C M H V Y K L
S I Y T M W I S C L N A F V C M Y T S *
1090 1100 1110 1120 1130 1140
Enhancer I
Liver Specific Regulatory Element (LSR)
---Binds NF1-
2C/TGT3b Site
AAACAGGCTTTCACTTTCTCGCCAACTTACAAGGCCTTTCTAAGTAAACAGTACAT
K Q A F T F S P T Y K A F L S K Q Y M N
N R L S L S R Q L T R P F * V N S T * T
T G F H F L A N L Q G L S K * T V H E P
1150 1160 1170 1180 1190 1200
Liver Specific Regulatory Element (LSR)
The EP Binding Site
<The E element Binding Site>
Binds NF1
eH-TF
> C/EBP Consensus:: CTGACGCAAC
C Consensus:RTTRCYNGGNRAY
AP-1 Consensus: TGAGTCA
CTTTACCCGTTGCTCGGCAACGCCTGGTCTGTGCCAAGTGTTTGCTGACGCAACCCCC
L Y P V A R Q R P G L C Q V F A D A T P
F T P L L G N G L V C A K C L L T Q P P
L P R C S A T A W S V P S V C * R N P H
1210 1220 1230 1240 1250 1260
(LSR)
X gene Promoter (X-P)
Binds NF1
ACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCATGCGTGGAACCTTTGTGGCTCCTCTG
Palindrome
T G W G L A I G H Q R M R G T F V A P L
L A G A W P * A I S A C V E P L W L L C
W L G L G H R P S A H A W N L C G S S A
1270 1280 1290 1300 1310 1320
(X-P)
CCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGCTCGCAGCCGGTCTGGAGCAAAG
P I H T A E L L A A C F A R S R S G A K
R S I L R N S * P L V L L A A G L E Q S
D P Y C G T P S R L F C S Q P V W S K A

```
     1330       1340       1350       1360       1370       1380
                                                    ->Start X
CTCATCGGAACTGACAATTCTGTCGTCCTCTCGCGGAAATATCATCGTTTCCATGGCTG
L  I  G  T  D  N  S  V  V  L  S  R  K  Y  T  S  F  P  W  L
 S  S  E  L  T  I  L  S  S  S  R  G  N  I  H  R  F  H  G  C
  H  R  N  *  Q  F  C  R  P  L  A  E  I  Y  I  V  S  M  A  A

     1390       1400       1410       1420       1430       1440
CTAGGCTGTACTGCCAACTGGATCCTTCGCGGGACGTCCTTTGTTTACGTCCCGTCGGCG
L  G  C  T  A  N  W  I  L  R  G  T  S  F  V  Y  V  P  S  A
 *  A  V  L  P  T  G  S  F  A  G  R  P  L  F  T  S  R  R  R
  R  L  Y  C  Q  L  D  P  S  R  D  V  L  C  L  R  P  V  G  A

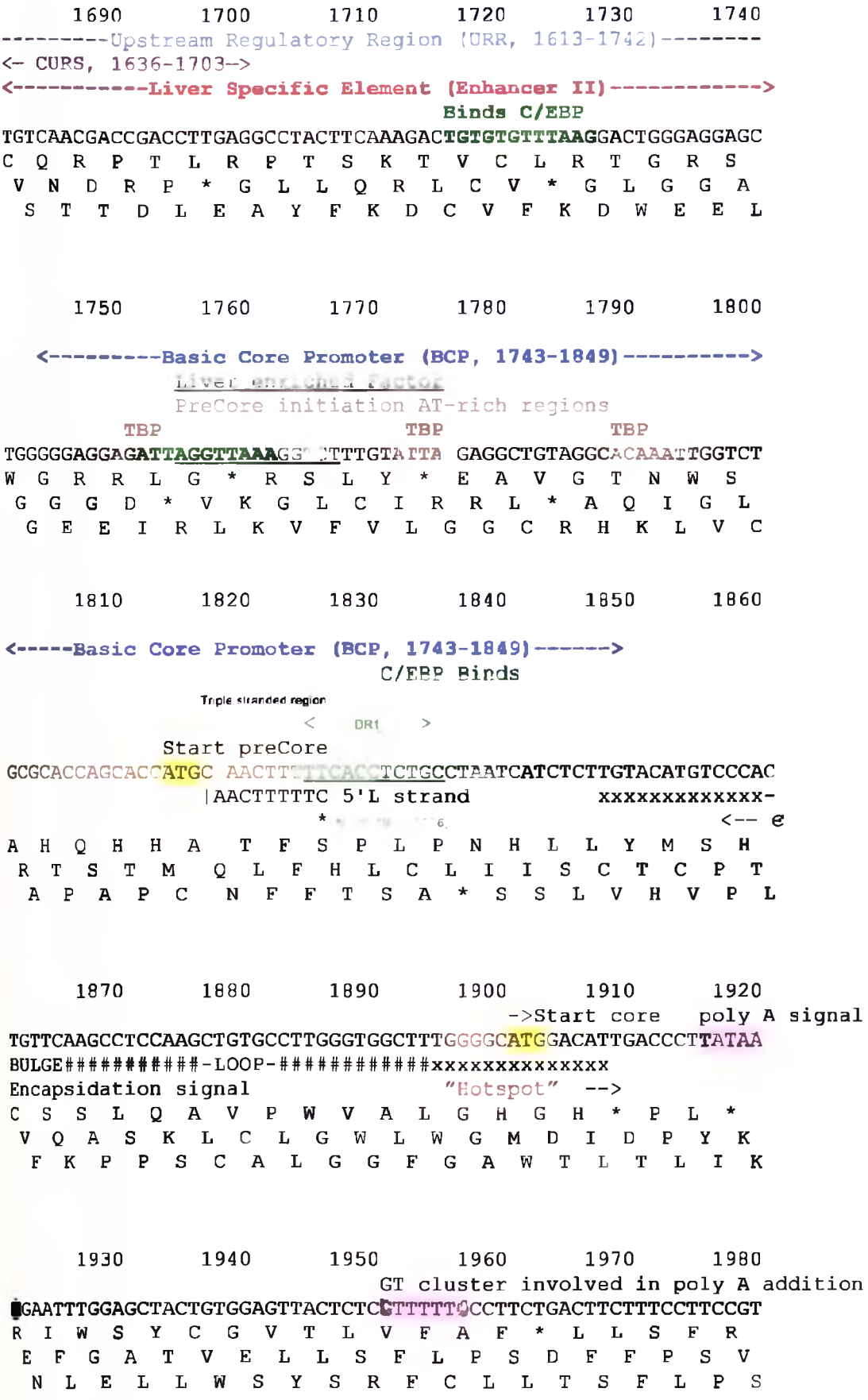
     1450       1460       1470       1480       1490       1500
CTGAATCCCGCGGACGACCCCTCTCGGGGCCGCTTGGGACTCTCTCGTCCCCTTCTCCGT
L  N  P  A  D  D  P  S  R  G  R  L  G  L  S  R  P  L  L  R
 *  I  P  R  T  T  P  L  G  A  A  W  D  S  L  V  P  F  S  V
  E  S  R  G  R  P  L  S  G  P  L  G  T  L  S  S  P  S  P  S

     1510       1520       1530       1540       1550       1560
CTGCCGTTCCAGCCGACCACGGGGCGCACCTCTCTTTACGCGGTCTCCCCGTCTGTGCCT
L  P  F  Q  P  T  T  G  R  T  S  L  Y  A  V  S  P  S  V  P
 C  R  S  S  R  P  R  G  A  P  L  F  T  R  S  P  R  L  C  L
  A  V  P  A  D  H  G  A  H  L  S  L  R  G  L  P  V  C  A  F

     1570       1580       1590       1600       1610       1620
                                                    <URR---
                          Start of the S strand

                          <    DR2   >
TCTCATCTGCCGGTCCGTGTGCACTTCGCTTCACCTCTGCACGTCGCATGGAGACCACCG
S  H  L  P  V  R  V  H  F  A  S  P  L  H  V  A  W  R  P  P
 L  I  C  R  S  V  C  T  S  L  H  L  C  T  L  H  G  D  H  R
  S  S  A  G  P  C  A  L  R  F  T  S  A  R  C  M  E  T  T  V

     1630       1640       1650       1660       1670       1680
---------Upstream Regulatory Region (URR, 1613-1742)--------
Negative Regulatory Element (NRE, 1611-1634)
               <-Core Upstream Regulatory Sequence (CURS, 1636-1703)
               End P
                 C/EBP     C/EBP Binds (2 sites)
TGAACGCCCATCAGATCCTGCCCAAGGTCTTACATAAGAGGACTCTTGGACTCCCAGCAA
*  T  P  I  R  S  C  P  R  S  Y  I  R  G  L  L  D  S  Q  Q
 E  R  P  S  D  P  A  Q  G  L  T  *  E  D  S  W  T  P  S  N
  N  A  H  Q  I  L  P  K  V  L  H  K  R  T  L  G  L  P  A  M
```


1690 1700 1710 1720 1730 1740
---------Upstream Regulatory Region (URR, 1613-1742)--------
<- CURS, 1636-1703->
<----------Liver Specific Element (Enhancer II)---------->
Binds C/EBP
TGTCAACGACCGACCTTGAGGCCTACTTCAAAGACTGTGTGTTTAAGGACTGGGAGGAGC
C Q R P T L R P T S K T V C L R T G R S
V N D R P * G L L Q R L C V * G L G G A
S T T D L E A Y F K D C V F K D W E E L
1750 1760 1770 1780 1790 1800
<---------Basic Core Promoter (BCP, 1743-1849)---------->
Liver enriched factor
PreCore initiation AT-rich regions
TBP TBP TBP
TGGGGGAGGAGATTAGGTTAAAGG TTTGTATTA GAGGCTGTAGGCACAAATTGGTCT
W G R R L G * R S L Y * E A V G T N W S
G G G D * V K G L C I R R L * A Q I G L
G E E I R L K V F V L G G C R H K L V C
1810 1820 1830 1840 1850 1860
<-----Basic Core Promoter (BCP, 1743-1849)------>
C/EBP Binds
Triple stranded region
< DR1 >
Start preCore
GCGCACCAGCACCATGC AACTTTTTCACCTCTGCCTAATCATCTCTTGTACATGTCCCAC
|AACTTTTTC 5'L strand xxxxxxxxxxxxx-
* <-- e
A H Q H H A T F S P L P N H L L Y M S H
R T S T M Q L F H L C L I I S C T C P T
A P A P C N F F T S A * S S L V H V P L
1870 1880 1890 1900 1910 1920
->Start core poly A signal
TGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAA
BULGE###########-LOOP-###########xxxxxxxxxxxxxx
Encapsidation signal "Hotspot" -->
C S S L Q A V P W V A L G H G H * P L *
V Q A S K L C L G W L W G M D I D P Y K
F K P P S C A L G G F G A W T L T L I K
1930 1940 1950 1960 1970 1980
GT cluster involved in poly A addition
GAATTTGGAGCTACTGTGGAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCCGT
R I W S Y C G V T L V F A F * L L S F R
E F G A T V E L L S F L P S D F F P S V
N L E L L W S Y S R F C L L T S F L P S

```
     1990       2000       2010       2020       2030       2040
CAGAGATCTCCTAGACACCGCCTCAGCTCTGTATCGAGAAGCCTTAGAGTCTCCTGAGCA
Q R S P R H R L S S V S R S L R V S * A
 R D L L D T A S A L Y R E A L E S P E H
  E I S * T P P Q L C I E K P * S L L S I

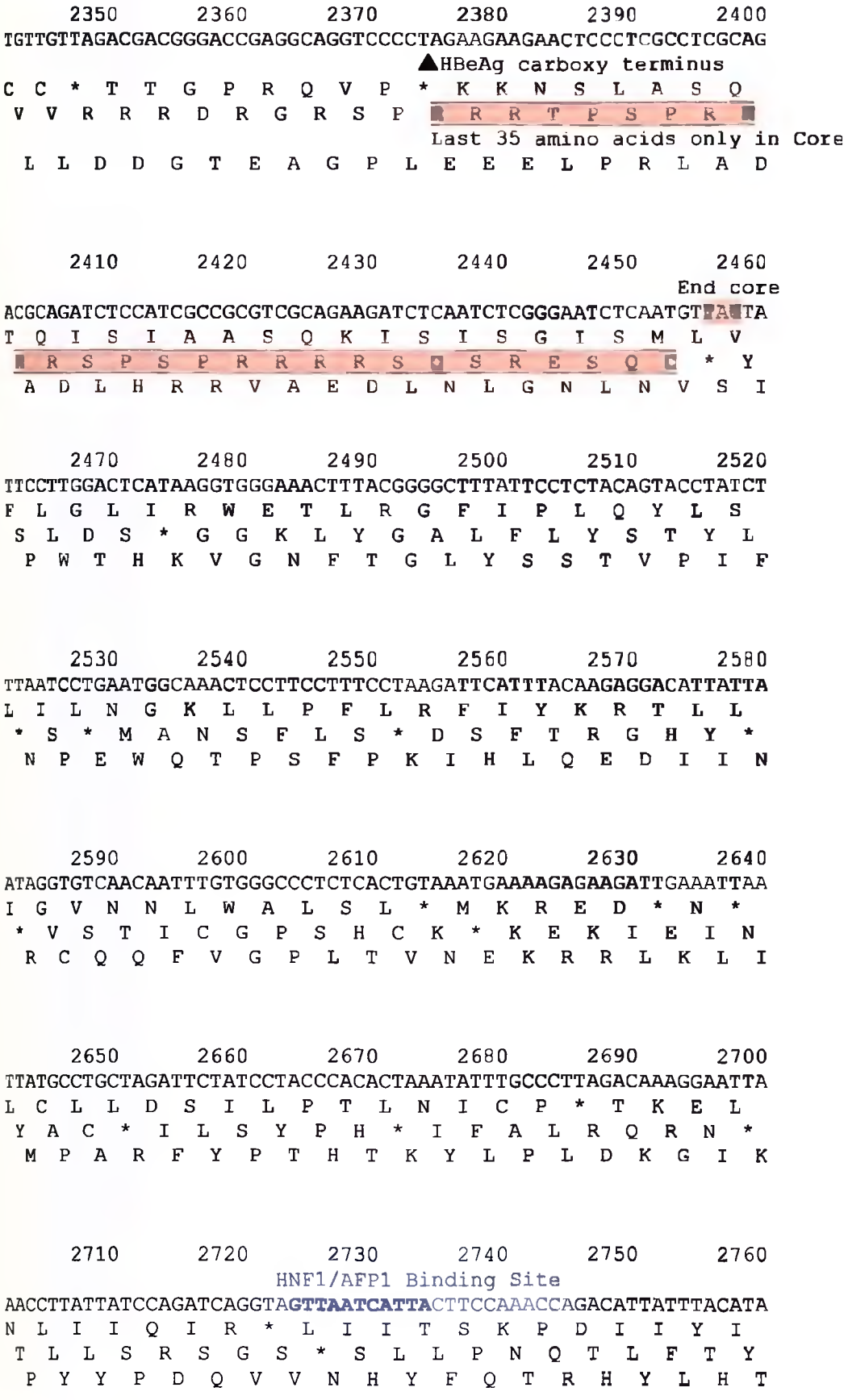
     2050       2060       2070       2080       2090       2100
TTGCTCACCTCACCATACTGCACTCAGGCAAGCCATTCTCTGCTGGGGGGAATTGATGAC
L L T S P Y C T Q A S H S L L G G I D D
 C S P H H T A L R Q A I L C W G E L M T
  A H L T I L H S G K P F S A G G N * * L

     2110       2120       2130       2140       2150       2160
                                       <-------- e1 epitope---
TCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCTAGGGATCTTGTAGTAAA
S S Y L G G * * F G R S S I * G S C S K
 L A T W V G N N L E D P A S R D L V V N
  * L P G W V I I W K I Q H L G I L * * I

     2170       2180       2190       2200       2210       2220
--->
TTATGTTAATACTAACGTGGGTTTAAAGATCAGGCAACTATTGTGGTTTCATATATCTTG
L C * Y * R G F K D Q A T I V V S Y I L
 Y V N T N V G L K I R Q L L W F H I S C
  M L I L T W V * R S G N Y C G F I Y L A

     2230       2240       2250       2260       2270       2280
CCTTACTTTTGGAAGAGAGACTGTACTTGAATATTTGGTCTCTTTCGGAGTGTGGATTCG
P Y F W K R D C T * I F G L F R S V D S
 L T F G R E T V L E Y L V S F G V W I R
  L L L E E R L Y L N I W S L S E C G F A

     2290       2300       2310       2320       2330       2340
      <------------e2 epitope--------->
                         ->Start P
CACTCCTCCAGCCTATAGACCACCAAATGCCCCTATCTTATCAACACTTCCGGAAACTAC
H S S S L * T T K C P Y L I N T S G N Y
 T P P A Y R P P N A P I L S T L P E T T
  L L Q P I D H Q M P L S Y Q H F R K L L
```


2350 2360 2370 2380 2390 2400
TGTTGTTAGACGACGGGACCGAGGCAGGTCCCCTAGAAGAAGAACTCCCTCGCCTCGCAG
▲HBeAg carboxy terminus
C C * T T G P R Q V P * K K N S L A S Q
V V R R R D R G R S P R R R T P S P R
Last 35 amino acids only in Core
L L D D G T E A G P L E E E L P R L A D
2410 2420 2430 2440 2450 2460
End core
ACGCAGATCTCCATCGCCGCGTCGCAGAAGATCTCAATCTCGGGAATCTCAATGTTAGTA
T Q I S I A A S Q K I S I S G I S M L V
R S P S P R R R R S S R E S Q C * Y
A D L H R R V A E D L N L G N L N V S I
2470 2480 2490 2500 2510 2520
TTCCTTGGACTCATAAGGTGGGAAACTTTACGGGGCTTTATTCCTCTACAGTACCTATCT
F L G L I R W E T L R G F I P L Q Y L S
S L D S * G G K L Y G A L F L Y S T Y L
P W T H K V G N F T G L Y S S T V P I F
2530 2540 2550 2560 2570 2580
TTAATCCTGAATGGCAAACTCCTTCCTTTCCTAAGATTCATTTACAAGAGGACATTATTA
L I L N G K L L P F L R F I Y K R T L L
* S * M A N S F L S * D S F T R G H Y *
N P E W Q T P S F P K I H L Q E D I I N
2590 2600 2610 2620 2630 2640
ATAGGTGTCAACAATTTGTGGGCCCTCTCACTGTAAATGAAAAGAGAAGATTGAAATTAA
I G V N N L W A L S L * M K R E D * N *
* V S T I C G P S H C K * K E K I E I N
R C Q Q F V G P L T V N E K R R L K L I
2650 2660 2670 2680 2690 2700
TTATGCCTGCTAGATTCTATCCTACCCACACTAAATATTTGCCCTTAGACAAAGGAATTA
L C L L D S I L P T L N I C P * T K E L
Y A C * I L S Y P H * I F A L R Q R N *
M P A R F Y P T H T K Y L P L D K G I K
2710 2720 2730 2740 2750 2760
HNF1/AFP1 Binding Site
AACCTTATTATCCAGATCAGGTAGTTAATCATTACTTCCAAACCAGACATTATTTACATA
N L I I Q I R * L I I T S K P D I I Y I
T L L S R S G S * S L L P N Q T L F T Y
P Y Y P D Q V V N H Y F Q T R H Y L H T

```
     2770       2780       2790       2800       2810       2820
                          "The" TATA BOX
CTCTTTGGAAGGCTGGTATTCTATATAAGCGGGAAACCACACGTAGCGCATCATTTTGCG
L F G R L V F Y I S G K P H V A H H F A
 S L E G W Y S I * A G N H T * R I I L R
  L W K A G I L Y K R E T T R S A S F C G

     2830       2840       2850       2860       2870       2880
                                      -----33-bp-deletion------
                                   preS1 Start
GGTCACCATATTCTTGGGAACAAGAGCTACAGCATGGGAGGTTGGTCATCAAAACCTCGC
G H H I L G N K S Y S M G G W S S K P R
 V T I F L G T R A T A W E V G H Q N L A
  S P Y S W E Q E L Q H G R L V I K T S Q

     2890       2900       2910       2920       2930       2940
AAAGGCATGGGGACGAATCTTTCTGTTCCCAATCCTCTGGGATTCTTTCCCGATCATCAG
K G M G T N L S V P N P L G F F P D H Q
 K A W G R I F L F P I L W D S F P I I S
  R H G D E S F C S Q S S G I L S R S S V

     2950       2960       2970       2980       2990       3000
TTGGACCCTGCATTCGGAGCCAACTCAAACAATCCAGATTGGGACTTCAACCCCGTCAAG
L D P A F G A N S N N P D W D F N P V K
 W T L H S E P T Q T I Q I G T S T P S R
  G P C I R S Q L K Q S R L G L Q P R Q G

     3010       3020       3030       3040       3050       3060
    NF1 Binding Site
GACGACTGGCCAGCAGCCAACCAAGTAGGAGTGGGAGCATTCGGGCCAAGGCTCACCCCT
D D W P A A N Q V G V G A F G P R L T P
 T T G Q Q P T K * E W E H S G Q G S P L
  R L A S S Q P S R S G S I R A K A H P S

     3070       3080       3090       3100       3110       3120
CCACACGGCGGTATTTTGGGGTGGAGCCCTCAGGCTCAGGGCATATTGACCACAGTGTCA
P H G G I L G W S P Q A Q G I L T T V S
 H T A V F W G G A L R L R A Y * P Q C Q
  T R R Y F G V E P S G S G H I D H S V N
```

```
     3130       3140       3150       3160       3170       3180
Activates PreS/S Promoter in liver cells but deactivates HELA cells
ACAATTCCTCCTCCTGCCTCCACCAATCGGCAGTCAGGAAGGCAGCCTACTCCCATCTCT
T  I  P  P  P  A  S  T  N  R  Q  S  G  R  Q  P  T  P  I  S
 Q  F  L  L  L  P  P  P  I  G  S  Q  E  G  S  L  L  P  S  L
  N  S  S  S  C  L  H  Q  S  A  V  R  K  A  A  Y  S  H  L  S


     3190       3200       3210       3220
            Enhancer Promoter-like element
            Binds AP-1/ CRE / SRE
                                  ->Start preS2
CCACCTCTAAGAGACAGTCATCCTCAGGCCATGCAGTGGAA 3'
P  P  L  R  D  S  H  P  Q  A  M  Q  W
 H  L  *  E  T  V  I  L  R  P  C  S  G
  T  S  K  R  Q  S  S  S  G  H  A  V  E
```

APPENDIX B

GENBANK AND LOCAL (SEQUENCED) PATIENT DETAILS

Table 25. GenBank and local (sequenced) patient details

	AUTHOR	Acc No.	REF	ID	bp	DIAGNOSIS	Bowyer *et al.*, 1997	Bowyer *et al.*, 2000	sequence
1	Kremsdorf *et al.*, 1996		A01		3221	HBsAg-negative persistent hepatitis	A04	A01	full genome
2	Estacio *et al.*, 1988	M57663	A02	HEADWZCG	3221	blood donor	A42	A02	full genome
3	Core sequence study, Chapter 10		A03	ES4	674	CAH (SA)			full core
4	Valliammai *et al.*, 1995	Z68141	A04		587	chronic infection (India)			full core
5	Bowyer *et al.*, 1997	U87740	A05	H510	846	acute HBV	A18		preS2 + HBsAg
6	Bowyer *et al.*, 1997	U87728	A06	E20	846	chronic carrier	A19		preS2 + HBsAg
7	Bowyer *et al.*, 1997	U87742	A07	7983	846	acute HBV	A20		preS2 + HBsAg
8	Bowyer *et al.*, 1997	U87731	A08	E17	846	chronic carrier	A23		preS2 + HBsAg
9	Bowyer *et al.*, 1997	U87733	A09	E30	846	chronic carrier	A25		preS2 + HBsAg
10	Bowyer *et al.*, 1997	U87852	A10	1149	846	acute HBV	A27		preS2 + HBsAg
11	Bowyer *et al.*, 1997	U87745	A11	1449	846	acute HBV	A29		preS2 + HBsAg
12	Bowyer *et al.*, 1997	U87743	A12	IR	795	acute HBV	A30		preS2 + HBsAg
13	Bowyer *et al.*, 1997	U87734	A13	E11	846	chronic HBV	A32		preS2 + HBsAg
14	Bowyer *et al.*, 1997	U87736	A14	E06	846	chronic HBV	A33		preS2 + HBsAg
15	Bowyer *et al.*, 1997	U87746	A15	7782	846	acute HBV	A40		preS2 + HBsAg
16	Chirara and Chetsanga, 1994	X69458	A16	HBVPRESS	1232	Zimbabwean variant	A45		full surface
17	Bowyer *et al.*, 1997	U87735	A17	A23	846	chronic carrier	A36		preS2 + HBsAg
18	GenBank direct submission	X51970	A01	HVHEPB HBV991	3221	HBV991	A07	A01	full genome
19	GenBank direct submission	E00010	A02		3221	non-transit virus		A02	full genome
20	Valenzuela *et al.*, 1979	X02763	A03	HBVADW2	3221	human sera	A44	A03	full genome

Table 25 continued overleaf

	AUTHOR	ACC No	REF	ID	bp	DIAGNOSIS	Bowyer *et al.*, 1997	Bowyer *et al.*, 2000	sequence
21	Preisler-Acams *et al.*, 1993	X70185	A04	HBVXCPS	3221		A08	A04	full genome
22	GenBank direct submission	Z72478	A05	HBVA317	3221	immunologically negative infection		A05	full genome
23	GenBank direct submission	Z35717	A06	HB614	3221	Polish isolate		A06	full genome
24	Ono *et al.*, 1983	V00866 J02201 X00715	A07	HBVADW ONOADW	3200	infected chimpanzee	A43	A07	full genome
25	GenBank direct submission	L13994	A08	HUMPRECX	3221	variant	A13	A08	full genome
26	Wands *et al.*, 1992	S50225	A09		3222	HBsAg negative persistent hepatitis - replication defective	A03	A09	full genome
27	Gunther *et al.*, 1996	L42364	A10	G2	525	renal transplant, deletion mutant, end stage disease			full core
28	Gunther *et al.*, 1996	L42361	A11	D2	621	renal transplant, deletion mutant, end stage disease			full core
29	Gunther *et al.*, 1996	L42363	A12	F5	582	renal transplant, deletion mutant, end stage disease			full core
30	Takahashi *et al.*, 1998	AB014370	A13	11D11HCC	3221	HCC			full genome
31	Cheng *et al.*, 1988	X07911	A14		645	adw2 patient			full core
32	GenBank direct submission	Z72479	A15		3216	immunologically negative infection			full genome
33	Norder *et al.*, 1992	X75669	A16	HHVBAMAM	681		A01		HBsAg
34	Lo *et al.*, 1986	M12346	A17	HPBSWADW	165		A02		preS2
35	Norder *et al.*, 1992	X75666	A18	HHVBA (Meg)	681	workshop no 89	A06		HBsAg
36	Rivkina *et al.* 1988	M21030	A19	HPBS	681	PLC/PRF/5	A09		HBsAg
37	Kidd-Ljunggren *et al.*, 1994		A20	LS	165	Philippine patient	A10		preS2
38	Kidd-Ljunggren *et al.*, 1994		A21	JOA	165	Polish patient	A11		preS2
39	Kidd-Ljunggren *et al.*, 1994		A22	MP	165	Polish patient	A12		preS2
40	Bowyer *et al.*, 1997	U87739	A23	1385	846	acute HBV	A14		preS2 + HBsAg
41	Bowyer *et al.*, 1997	U87725	A24	A24	846	chronic HBV	A15		preS2 + HBsAg
42	Bowyer *et al.*, 1997	U87726	A25	P08	846	acute HBV	A16		preS2 + HBsAg
43	Bowyer *et al.*, 1997	U87727	A26	E14	846	chronic HBV	A17		preS2 + HBsAg
44	Bowyer *et al.*, 1997	U87729	A27	P03	846	chronic HBV	A21		preS2 + HBsAg
45	Bowyer *et al.*, 1997	U87730	A28	O04	846	chronic HBV	A22		preS2 + HBsAg

Table 25 continued overleaf

	AUTHOR	Acc. No	REF	ID	bp	DIAGNOSIS	Bowyer *et al.*, 1997	Bowyer *et al.*, 2000	sequence
46	Bowyer *et al.*, 1997	U87732	A29	C01	846	chronic HBV	A24		preS2 + HBsAg
47	Bowyer *et al.*, 1997	U87741	A30	1019	846	acute HBV	A26		preS2 + HBsAg
48	Bowyer *et al.*, 1997	U88094	A31	C02	846	chronic HBV	A31		preS2 + HBsAg
49	Bowyer *et al.*, 1997	U87844	A32	8017	846	acute HBV	A28		preS2 + HBsAg
50	Bowyer *et al.*, 1997	U88096	A33	S01	165	chronic HBV	A38		pres2
51	Bowyer *et al.*, 1997	U88095	A34	E34	165	chronic HBV	A34		pres2
52	Ting and Chang, 1987	M54898	A35	HPBVSAG2	1371	chronic HBV	A41		partial preS2+ HBsAg+
53	Pult *et al.*, 1997	X98077	B'01	HBVCGWITY (wild type pF-6)	3215	transplant transmitted fulminant hepatitis		B'01	full genome
54	Okamoto *et al.*, 1988	D00330	B'02	HPBADW2/ pODW282	3215	asymptomatic carrier	B04	B'02	full genome
55	Alexopoulou *et al.*, 1996	X97851	B'03	HBVP6CSX/ patient 6	3215	persistent infection		B'03	full genome
56	Alexopoulou *et al.*, 1996	X97850	B'04	HBVP4CSX/ patient 4	3215	fulminant		B'04	full genome
57	Hannoun *et al.*, 2000	AF121244	B'05	HBV/ 93-15122	3215	longterm infection (family 2, child 1)			full genome
58	Hannoun *et al.*, 2000	AF121245	B'06	HBV/ 92-16091	3215	longterm infection (family 2, child 2)			full genome
59	Hannoun *et al.*, 2000	AF121243	B'07	HBV/ 95-14611	3215	longterm infection (family 2,mother)			full genome
60	Hannoun *et al.*, 2000	AF121246	B'08	HBV/ 93-15122	3215	longterm infection (family 3, child 3)			full genome
61	Hannoun *et al.*, 2000	AF121249	B'09	HBV/ 98-4014	3215	longterm infection (family 3, child 2)			full genome
62	Hannoun *et al.*, 2000	AF121247	B'10	HBV/ 94-719	3215	longterm infection (family 3, mother)			full genome
63	Hannoun *et al.*, 2000	AF121251	B'11	HBV/ 97-9873	3215	longterm infection (family 3, child 4)			full genome
64	Hannoun *et al.*, 2000	AF121248	B'12	HBV/ 139-1	3215	longterm infection (family 3, child 1)			full genome
65	Hannoun *et al.*, 2000	AF121250	B'13	HBV/ 97-6395	3215	longterm infection (family 3, child 3)			full genome
66	Pult *et al.*, 1997	X98074	B'14	pF-3	3024	transplant transmitted fulminant hepatitis (new HNF1 site)			full genome

Table 25 continued overleaf

	AUTHOR	Acc. No.	REF	ID	bp	DIAGNOSIS	Bowyer *et al.*, 1997	Bowyer *et al.*, 2000	sequence
67	Pult *et al.*, 1997	X98076	B15	pF-5	2761	transplant transmitted fulminant hepatitis (new HNF1 site)			full genome (338 bp deletion)
68	Pult *et al.*, 1997	X98075	B16	pF-4	3297	transplant transmitted fulminant hepatitis (new HNF1 site)			full genome (198 bp insertion)
69	Pult *et al.*, 1997	X98072	B17	pF-1	3100	transplant transmitted fulminant hepatitis (new HNF1 site)			full genome (with insertion)
70	Pult *et al.*, 1997	X98073	B18	pF-2	3100	transplant transmitted fulminant hepatitis (new HNF1 site)			full genome (with insertion)
71	Norder *et al.*, 1992	X75660	B01	HVPRS	681		B07		HBsAg
72	Koseki *et al.*, 1999	AB010289	B02	Yamagata 1	3215	fulminant hepatitis			full genome
73	Takahashi *et al.*, 1998	AB014366	B03	07D07HCC	3227	HCC			full genome
74	Koseki *et al.*, 1999	AB010292	B04	Yamagata 4	3215	fulminant hepatitis			full genome
75	Asahina *et al.*, 1996	D50522	B05	HBV ASA-FH2	3215	haemodialysis patient epidemic fatal fulminant hepatitis		B05	full genome
76	Asahina *et al.*, 1996	D50521	B06	HBV ASA FH1	3215	haemodialysis patient - epicemic fatal fulminant hepatitis		B06	full genome
77	Horikita *et al.*, 1994	D23677	B07	HPBA1HKK2 A1-HBVKK2	3207	asymptomatic carrier		B07	full genome
78	Horikita *et al.*, 1994	D23678	B08	HPBA2HYS2 A2-HBVYS2	3215	asymptomatic carrier		B08	full genome
79	Horikita *et al.*, 1994	D23679	B09	HPBA3HMS2 A3-HBVHS2	3215	asymptomatic carrier		B09	full genome
80	Okamoto *et al.*, 1988	D00329	B10	HPBADW1 pJDW233	3215	asymptomatic carrier	B02	B10	full genome
81	Koseki *et al.*, 1999	AB010291	B11	Yamagata-3	3215	fulminant hepatitis			full genome
82	Koseki *et al.*, 1999	AB010290	B12	Yamagata-2	3215	fulminant hepatitis			full genome
83	Bowyer *et al.*, 1997	U87747	B13	WONG	840	acute HBV	B05		pres2 + Surface
84	Lauder *et al.*, 1993		B14	ISWARI	165		B06		pres2
85	Okamoto *et al.*, 1988	D00331	B*1	HPBADW3 pIDW420	3215	asymptomatic carrier	B01	B*11	full genome
86	Sastrosoewignjo *et al.*, 1987	M54923	B*2	HPBADWZ pMND122	3215		B03	B*12	full genome

Table 25 continued overleaf

	AUTHOR	Acc. No.	REF	ID	bp	DIAGNOSIS	Bowyer *et al.*, 1997	Bowyer *et al.*, 2000	sequence
	Norder *et al.*, 1994	X75656	C01	HHVCCHA	3215	French Polynesian carrier adrq-	C24	Cq 01	full genome
	Norder *et al.*, 1994	X75665	C02	HHVBC HMA	3215	New Caledonian carrier adrq-	C20	Cq- 02	full genome
	Kobayashi and Koike, 1984	D00630	C03	HPBADRC	3215	CPH		C03	full genome
	Horikita *et al.*, 1994	D23684	C04	HPBC6T588 C6-TKB588	3215	CAH		C04	full genome
	Horikita *et al.*, 1994	D23681	C05	C4- HBVST2	3215	fulminant		C05	full genome
	Horikita *et al.*, 1994	D23680	C06	R4 HBVST1	3194	CPH		C06	full genome
	Fujiyama *et al.*, 1983	X01587	C07	HBVADR4	3214		C12	C07	full genome
	Renbac *et al.*, 1987	M38454	C08	HPBADR1CG pADR1	3215		C02	C08	full genome
	Ono *et al.*, 1983	V00867	C09	HBVADR ONOADR	3188	CPH	C09	C09	full genome
	Okamoto *et al.*, 1986	X04615	C10	HEHBVAYR	3215		C16	C10	full genome
	Kobayashi *et al.*, 1984	M12906	C11	HPBADRA	3215		C11	C11	full genome
	Mukaide *et al.*, 1992	D12980	C12	HPBCG SRADR	3215	CAH	C08	C12	full genome
	Uchida *et al.*,1995	D16667	C13	patient H2	3207	core promoter mutants with truncated X genes		C13	full genome
	Uchida *et al.*,1995	D16666	C14	isolate E88	3192	core promoter mutants with truncated X genes		C14	full genome
	Uchida *et al.*,1995	D50489	C15	isolate A11	3192	core promoter mutants with truncated X genes		C15	full genome
	Horikita *et al.*, 1994	D23682	C16	HPBB5HK01/ B5-HBVK01	3215	CPH		C16	full genome
	Horikita *et al.*, 1994	D23683	C17	HPBC5HK02/ C5-HBVK02	3215	CAH		C17	full genome
	Horikita *et al.*, 1994	D23683	C17	HPBC5HK02/ C5-HBVK02	3215	CAH		C17	full genome
	Ogata *et al.*, 1993	L08805	C18	HPBETNC	3212	fulminant (lab infected chimpanzee)	C10	C18	full genome
	Takahashi *et al.*, 1998	AB014387	C19	28Y10HCC	3170	HCC			full genome
	Takahashi *et al.*, 1998	AB014376	C20	17T02HCC	3221	HCC			full genome
	Ashina *et al.*, 1996	D50520	C21	HBV-ASA EX4	3215	acute exacerbation			full genome
	Takahashi *et al.*, 1998	AB014363	C22	04D04HCC	3215	HCC			full genome

Table 25 continued overleaf

	AUTHOR	Acc. No	REF	ID	bp	DIAGNOSIS	Bowyer *et al.*, 1997	Bowyer *et al.*, 2000	sequence
109	Asahina *et al.*, 1996	D50519	C23	HBV-ASA-EX3	3215	acute exacerbation			full genome
110	Asahina *et al.*, 1996	D50517	C24	HBV-ASA-EX1	3215	acute exacerbation			full genome
111	Asahina *et al.*, 1996	D50518	C25	HBV-ASA-EX2	3215	acute exacerbation			full genome
112	Takahashi *et al.*, 1998	AB014375	C26	16T01HCC	3170	HCC			full genome
113	Takahashi *et al.*, 1998	AB014365	C27	06D06HCC	3215	HCC			full genome
114	Takahashi *et al.*, 1998	AB014378	C28	16Y01HCC	3215	HCC			full genome
115	Takahashi *et al.*, 1998	AB014377	C29	16T03HCC	3215	HCC			full genome
116	Takahashi *et al.*, 1998	AB014394	C30	16Y17HCC	3215	HCC			full genome
117	Takahashi *et al.*, 1998	AB014367	C31	08D08HCC	3215	HCC			full genome
118	Takahashi *et al.*, 1998	AB014373	C32	14D14HCC	3179	HCC			full genome
119	Takahashi *et al.*, 1998	AB014374	C33	15D15HCC	3215	HCC			full genome
120	Takahashi *et al.*, 1998	AB014390	C34	16Y13HCC	3176	HCC			full genome
121	Takahashi *et al.*, 1998	AB014372	C35	13D13HCC	3197	HCC			full genome
122	Takahashi *et al.*, 1998	AB014368	C36	09D09HCC	3167	HCC			full genome
123	Wakita *et al.*, 1991	S71628	C37		639	chronic carrier			full core
124	Kim *et al.*, 1988	M38636	C38	HPBCGADR	3213	type adr	C17		full genome
125	Rho *et al.*, 1989	X14193	C39	HBVADRM	3213	adr mutant	C18		full genome
126	Kim *et al.*, 1988	M38594	C40	adr-k	1257				full core
127	Takahashi *et al.*, 1998	AB014395	C41	36Y18HCC	3119	HCC			full genome
128	Takahashi *et al.*, 1998	AB014393	C42	34Y16HCC	3215	HCC			full genome
129	Takahashi *et al.*, 1998	AB014397	C43	38Y20HCC	3185	HCC			full genome
130	Takahashi *et al.*, 1998	AB014379	C44	20Y02HCC	3215	HCC			full genome
131	Takahashi *et al.*, 1998	AB014383	C45	24Y06HCC	3215	HCC			full genome
132	Takahashi *et al.*, 1998	AB014380	C46	21Y03HCC	3215	HCC			full genome
133	Takahashi *et al.*, 1998	AB014392	C47	33Y15HCC	3215	HCC			full genome
134	Takahashi *et al.*, 1998	AB014382	C48	23Y05HCC	3215	HCC			full genome
135	Takahashi *et al.*, 1998	AB014384	C49	25Y07HCC	3215	HCC			full genome
136	Takahashi *et al.*, 1998	AB014399	C50	40Y22HCC	3215	HCC			full genome
137	Takahashi *et al.*, 1998	AB014371	C51	11D11HCC	3221	HCC			full genome

Table 25 continued overleaf

	AUTHOR	Acc. No	REF	ID	bp	DIAGNOSIS	Bowyer et al., 1997	Bowyer et al., 2000	sequence
138	Takahashi *et al.*, 1998	AB014369	C52	10D10HCC	3215	HCC			full genome
139	Takahashi *et al.*, 1998	AB014391	C53	32Y14HCC	3215	HCC			full genome
140	Takahashi *et al.*, 1998	AB014396	C54	37Y19HCC	3215	HCC			full genome
141	Takahashi *et al.*, 1998	AB014388	C55	29Y11HCC	3161	HCC			full genome
142	Takahashi *et al.*, 1998	AB014381	C56	22Y04HCC	3215	HCC			full genome
143	Takahashi *et al.*, 1998	AB014385	C57	26Y08HCC	3215	HCC			full genome
144	Takahashi *et al.*, 1998	AB014364	C58	05D05HCC	3104	HCC			full genome
145	Mukaide *et al.*, 1992	D16665	C59	Patient A4	3193				full genome
146	Muriyama *et al.*, 1994	D28880	C60	1992-085-9434	3215	HBeAg-negative isolate			full genome
147	Takahashi *et al.*, 1998	AB014389	C61	30Y12HCC	3168	HCC			full genome
148	Takahashi *et al.*, 1998	AB014362	C62	03D03HCC	3215	HCC			full genome
149	Takahashi *et al.*, 1998	AB014361	C63	02D02HCC	3167	HCC			full genome
150	Takahashi *et al.*, 1998	AB014386	C64	27Y09HCC	3014	HCC			full genome
151	Takahashi *et al.*, 1998	AB014398	C65	39Y21HCC	3167	HCC			full genome
152	Core sequence study, Chapter 10		C66	ES1	692	CAH (SA)			full core
153	Takahashi *et al.*, 1998	AB014360	C67	01D01HCC	3215	HCC			full genome
154	Core sequence study, Chapter 10		C68	ES3	696	CAH (SA)			full core
155	Lauder *et al.*, 1993		C69	OKAADW	165		C01		preS2
156	Kidd-Ljunggren *et al.*, 1994		C70	CP	165		C03		preS2
157	Ohno *et al.*, 1993	S62754	C71	B12	1203	defective short variant of S (chronic HBV)	C04		full surface
158	Okamoto et al., 1987	M17688	C72	OKAADYR	678	blood donor	C05		HBsAg
159	Norder et al., 1992	X75792	C73	HHVBCS O11 workshop no. 30	681		C06		HBsAg
160	Kidd-Ljunggren *et al.*, 1994		C74	DD	165		C07		preS2
161	Takahashi *et al.*, 1998	M23805	C75	HPBHBSAGA pHBR3	681		C13		HBsAg
162	Takahashi *et al.*, 1998	M23806	C76	HPBHBSAGR pHBR14	681		C14		HBsAg

Table 25 continued overleaf

	AUTHOR	Acc. No.	REF	ID	bp	DIAGNOSIS	Bowyer et al., 1997	Bowyer et al., 2000	sequence
163	Takahashi *et al.*, 1998	M23808	C77	HPBHBSAGD pHBR31	681		C15		HBsAg
164	Kidd-Ljunggren *et al.*, 1994		C78	KRJ	165		C19		preS2
165	Kidd-Ljunggren *et al.*, 1994		C79	491	165		C21		preS2
166	Kidd-Ljunggren *et al.*, 1994		C80	651	165		C22		preS2
167	Kidd-Ljunggren *et al.*, 1994		C81	029	165		C23		preS2
168	Bowyer *et al.*, 1997	U87748	C82	LEE	846		C25		preS2 + HBsAg
169	Kidd-Ljunggren *et al.*, 1994		C83	AF	165		C26		preS2
170	GenBank direct submission	Z35716	D01	HB321	3182	polish isolate ayw4		D01	full genome
171	Priesler-Adams, 1993	X72702	D02	HBVORFS C1005	3182	no evidence of infection	D04	D02	full genome
172	Alexopoulou *et al.*, 1996	X97848	D03	HBVP2CSX patient 2	3182	fulminant		D03	full genome
173	Alexopoulou *et al.*, 1996	X97849	D04	HBVP3CSX patient 3	3182	fulminant		D04	full genome
174	Karyiannis *et al.*, 1995	X80925	D05	patient 6	3182	CAH		D05	full genome
175	Karyiannis *et al.*, 1995	X80926	D06	patient 5	3182	CPH (anti-e + carrier)		D06	full genome
176	Karyiannis *et al.*, 1995	X80924	D07	patient 4	3182	fulminant		D07	full genome
177	Tong *et al.*, 1990	M32138	D08	HPBHBVAA/ HBV alpha1	3182	CAH (anti-e + carrier)	D25	D08	full genome
178	Lai *et al.*, 1991	X59795	D09	HBVAYWMCG patient MG	3161	CAH (anti-e + carrier)	D10	D09	full genome
179	Haseçawa *et al.*, 1994	L27106	D10		3185	fulminant (mutant)		D10	full genome
180	Bichko *et al.*, 1985	X02496	D11	XXHEPAV	3182	HBV from Latvia (variant)	D13	D11	full genome
181	GenBank direct submission	Y07587	D12		3182	immunotolerant carrier (variant)		D12	full genome
182	Galibert *et al.*, 1979	J02203 V01460	D13	HPBAYW pHB320	3182	ayw3 France	D01	Dm13	full genome
183	GenBank direct submission	U95551	D14		3182	genome from hepatoblastoma cell line 2.2.15		Dm14	full genome
184	GenBank direct submission	X65257	D15	HBVAYWC patient C	3182	HBsAg negative persistent hepatitis	D09	Dm15	full genome
185	GenBank direct submission	X65258	D16	HBVAYWCI patient CI	3182	HBsAg negative persistent hepatitis	D07	Dm16	full genome
186	GenBank direct submission	X65259	D17	HBVAYWE patient E	3182	HBsAg negative persistent hepatitis	D08	Dm17	full genome

Table 25 continued overleaf

	AUTHOR	Acc No.	REF	ID	bp	DIAGNOSIS	Bowyer et al., 1997	Bowyer et al., 2000	sequence
187	GenBank direct submission	X68292	D18	HBVDNA	3182	HBsAg-negative persistent hepatitis	D11	Dm18	full genome
188	Pasek *et al.*, 1979	J02202	D19	HPBADYW	2743	Pasek adyw	D18		full genome
189	Hannoun *et al.*, 2000	AF121240	D20	HBV/94-11066	3182	longterm infection (family 1, child 1)			full genome
190	Hannoun *et al.*, 2000	AF121239	D21	HBV/93-1303	3182	longterm infection (family 1, mother)			full genome
191	Hannoun *et al.*, 2000	AF121241	D22	HBV/95-1758	3182	longterm infection (family 1, child 2)			full genome
192	Hannoun *et al.*, 2000	AF121242	D23	HBV/98-1218	3182	longterm infection (family 1, child 3)			full genome
193	Karyiannis *et al.*, 1995	X80928	D24	patient 1	639	HBeAg-negative fulminant hepatitis			full core
194	Karyiannis *et al.*, 1995	X80927	D25	patient 2	639	HBeAg-negative fulminant hepatitis			full core
195	GenBank direct submission	X85312	D33	HBVXPCC58 Licheri	712				full core
	GenBank direct submission	X85313	D32	HBVXPCC59 Licheri 1	712				full core
	GenBank direct submission	X85314	D26	HBVXPCC60 Licheri 2	712				full core
	GenBank direct submission	X85315	D30	HBVXPCC61 Licheri 3	712				full core
196	GenBank direct submission	X85317	D27	HBVXPCC63 Giordo 2	712				full core
	GenBank direct submission	X85316	D28	HBVXPCC62 Giordo	712				full core
197	GenBank direct submission	X85300	D29	HBVXPCC46 Sanna	712				full core
198	Core sequence study, Chapter 10		D31	ES2	651	CAH (SA)			full core
199	GenBank direct submission	X85319	D34	HBVXPCC65 Galistru 1	712				full core
200	Norder *et al.*, 1992	X75668	D35	HHVBDS Mau workshop 85	681		D02		HBsAg
201	Kidd-Ljunggren *et al.*, 1994		D36	PC	165		D05		preS2
202	Kidd-Ljunggren *et al.*, 1994		D37	NB	165		D06		preS2

Table 25 continued overleaf

	AUTHOR	Acc No.	REF	ID	bp	DIAGNOSIS	Bowyer *et al.*, 1997	Bowyer *et al.*, 2000	sequence
203	Kidd-Ljunggren *et al.*, 1994		D38	TY5	165		D12		preS2
204	Kidd-Ljunggren *et al.*, 1994		D39	376	165		D14		preS2
205	Kidd-Ljunggren *et al.*, 1994		D40	567	165		D15		preS2
206	Kidd-Ljunggren *et al.*, 1994		D41	FEMI	165		D16		preS2
207	Pumpen *et al.*, 1984	M12393	D42	HPBHBSAG	829		D17		full surface
208	Kidd-Ljunggren *et al.*, 1994		D43	MAJ	165		D19		preS2
209	Bowyer *et al.*, 1997	U87737	D44	E08	846	chronic carrier	D20		preS2 + HBsAg
210	Bowyer *et al.*, 1997	U87738	D45	S09	846	chronic carrier	D21		preS2 + HBsAg
211	Bowyer *et al.*, 1997	U87851	D46	P67	846		D22		preS2 + HBsAg
212	Kidd-Ljunggren *et al.*, 1994		D47	AN	165		D26		preS2
213	Norder *et al.*, 1992	X75662	D48	HHVBDS	681		D03		HBsAg
214	Norder *et al.*, 1994	X75664	E01	HHVBE4 Kou	3212	Senegalese carrier	E04	E01	full genome
215	Norder *et al.*, 1994	X75657	E02	HHVBBAS	3212	West African carrier	E06	E02	full genome
216	Core sequence study, Chapter 10		E03	ES5	3212	CAH (SA)			full genome
217	Core sequence study, Chapter 10		E04	JB2	654	CAH (Namibia)			full core
218	Core sequence study, Chapter 10		E05	JB4	681	CAH (Namibia)			full core
219	Core sequence study, Chapter 10		E06	JB5	692	CAH (Namibia)			full core
220	Core sequence study, Chapter 10		E07	JB11	676	CAH (Namibia)			full core
221	Core sequence study, Chapter 10		E08	JB13	657	CAH (Namibia)			full core
222	Core sequence study, Chapter 10		E09	JB16	674	CAH (Namibia)			full core
223	Core sequence study, Chapter 10		E10	JB1	653	CAH (Namibia)			full core
224	Core sequence study, Chapter 10		E11	JB17	675	CAH (Namibia)			full core
225	Core sequence study, Chapter 10		E12	JB7	656	CAH (Namibia)			full core

Table 25 continued overleaf

	AUTHOR	Acc. No.	REF	ID	bp	DIAGNOSIS	Bowyer *et al.*, 1997	Bowyer *et al.*, 2000	sequence
226	Kidd-Ljunggren *et al.*, 1994		E13	YA	165		E01		preS2
227	Kidd-Ljunggren *et al.*, 1994		E14	JAM	165		E02		preS2
228	GenBank direct submission	L24071	E15	HPBVAR	681	variant in immunized child	E03		HBsAg
229	Naumann *et al.*, 1993	X69798	F01	HBVADW4A	3215	Brazilian blood donor	F01	F01	full genome
230	Norder *et al.*, 1994	X75663	F02	HHVBF 9203	3215	Colombian carrier	F02	F02	full genome
231	Norder *et al.*, 1994	X75658	F03	HHVBFFOU	3215	French carrier	F03	F03	full genome
232	Norder *et al.*, 1994	X75661	F04	HHVBFS	681	French carrier	F04		full genome
233	Stuyver *et al.*, 2000	AF160501	G01	IG29227	3248	chronic carrier			full genome
234	Tran *et al.*, 1991	M74502 M74499	G02	HPBCCB1MUT HPBSSB1MUT	675 1203	chronic HBV infection	 A05		full core full surface
235	Tran *et al.*, 1991	M74500	G03	HPBSSB1REC	1203	chronic HBV infection	A46		full surface
236	Takahashi *et al.*, 1998	AB032432	Ch01	ChHBV-Ch256	3182	chimpanzee wi			full genome
237	Hu *et al.*, 2000	AF222323	Ch02	CH926	3182	chimpanzee wi			full genome
238	Hu *et al.*, 2000	AF222322	Ch03	CH109	3182	chimpanzee wi			full genome
239	Vaudin *et al.*, 1988	D00220	Ch04	HPBVCG/LSH	3182	chimpanzee wi	E05	E03	full genome
240	Takahashi *et al.*, 1998	AB032433	Ch05	ChHBV-Ch258	3182	chimpanzee wi			full genome

Acc. No. = GenBank accession number
CAH = chronic active hepatitis
CPH = chronic persistent hepatitis

APPENDIX C

COMPUTER PROGRAMMES TO DETERMINE THE "HOTSPOTS" ON THE HBV GENOME

C.1 Reading data into the database

C.1.1 INPUT.PRG

*Programme INPUT.PRG, reads sequence data into the database file BASICDB.DBF

*Multiple fields A-N are required since the maximum field width is 240

```
USE BASICDB.dbf
STORE "X51970" TO THENAME
REPLACE NAME WITH THENAME
STORE "A01" TO THEREF
REPLACE REF WITH THEREF
STORE "[insert first 240 bp of sequence data here]" TO TEMP
REPLACE A WITH TEMP
STORE "[insert 241-480 bp of sequence data here]" TO TEMP
REPLACE B WITH TEMP
STORE "[insert 481-720 bp of sequence data here]" TO TEMP
REPLACE C WITH TEMP
STORE "[insert 721-960 bp of sequence data here]" TO TEMP
REPLACE D WITH TEMP
STORE "[insert 961-1200 bp of sequence data here]" TO TEMP
REPLACE E WITH TEMP
STORE "[insert 1201-1440 bp of sequence data here]" TO TEMP
REPLACE F WITH TEMP
STORE "[insert 1441-1680 bp of sequence data here]" TO TEMP
REPLACE G WITH TEMP
STORE "[insert 1681-1920 bp of sequence data here]" TO TEMP
```

```
REPLACE H WITH TEMP
STORE "[insert 1921-2160 bp of sequence data here]" TO TEMP
REPLACE I WITH TEMP
STORE "[insert 2161-2400 bp of sequence data here]" TO TEMP
REPLACE J WITH TEMP
STORE "[insert 2401-2640 bp of sequence data here]" TO TEMP
REPLACE K WITH TEMP
STORE "[insert 2641-2880 bp of sequence data here]" TO TEMP
REPLACE L WITH TEMP
STORE "[insert 2881-3120 bp of sequence data here]" TO TEMP
REPLACE M WITH TEMP
STORE "[insert 3121-3221 bp of sequence data here]" TO TEMP
REPLACE N WITH TEMP
SKIP 1
USE
```

C.1.2 Structure of the nucleotide sequence database BASICDB.DBF

Field	Fieldname	Type	Width	Dec
1	NAME	Character	10	
2	REF	Character	4	
3-16	A-N	Character	240	
17	TEMP	Character	240	

C.2 Translating a database of sequence data and storage of the predicted amino acid sequences

The file TRANSLAT.DBF used in the following three programmes is a look-up database file containing the various sequence codes and their protein equivalents in single letter format.

C.2.1 RF1.PRG

```
*RF1.PRG translates BASICDB.DBF in RF1 and stores data in PROT1.DBF
STORE 1 TO START
STORE 240 TO Y
STORE "ABCDEFGHIJKLMNA" TO FLDSTORE
```

```
STORE SUBSTR(FLDSTORE,START,1) TO THEFIELD
DO WHILE START<15
IF START=14
Y=101
ELSE
Y=240
ENDIF
STORE 1 TO Z
SELECT 1
USE BASICDB.DBF
GO TOP
STORE RECCOUNT() TO MAX
REPLACE ALL BASICDB->TEMP WITH SUBSTR(&THEFIELD,1,Y)
DO WHILE Z<(MAX+1)
STORE 1 TO X
STORE "" TO THESTORE
STORE " " TO THENAME
SELECT 1
USE BASICDB.DBF
GO Z
DO WHILE X<240
STORE " " TO THEAA
STORE " " TO THECODE
SELECT 1
STORE NAME TO THENAME
STORE SUBSTR(TEMP,X,3) TO THECODE
SELECT 2
USE TRANSLAT.DBF
IF SUBSTR(THECODE,1,1)="N".OR.SUBSTR(THECODE,2,1)="N".OR.SUBSTR
(THECODE,3,1)="N".OR.SUBSTR(THECODE,1,1)="O".OR.SUBSTR(THECODE,2,1)
="0".OR.SSUBSTR(THECODE,3,1)="O"
STORE "X" TO THEAA
ELSE
```

```
LOCATE FOR CODE=THECODE
STORE AA TO THEAA
ENDIF
THESTORE=THESTORE+THEAA
X=X+3
ENDDO
REPLACE FSTORE WITH THESTORE
SELECT 3
USE PROT1.DBF
GO Z
REPLACE PROT1->&THEFIELD WITH B->FSTORE
REPLACE NAME WITH THENAME
Z=Z+1
ENDDO
START=START+1
STORE SUBSTR(FLDSTORE,START,1) TO THEFIELD
ENDDO
SELECT 3
USE PROT1.DBF
GO TOP
*COMBINE THE FIELDS A-N INTO 240 CHARACTER FIELDS CALLED AA1-AA5
GO TOP
REPLACE ALL AA1 WITH A+B+C
REPLACE ALL AA2 WITH D+E+F
REPLACE ALL AA3 WITH G+H+I
REPLACE ALL AA4 WITH J+K+L
REPLACE ALL AA5 WITH M+N
USE
```

C.2.2 RF2.PRG

```
*RF2.PRG translates BASICDB.DBF in RF2 and stores data in PROT2.DBF
STORE 1 TO START
STORE 239 TO Y
```

```
STORE "ABCDEFGHIJKLMNA" TO FLDSTORE
STORE SUBSTR(FLDSTORE,START,1) TO THEFIELD
STORE SUBSTR(FLDSTORE,START+1,1) TO NXTFIELD
DO WHILE START<15
STORE 1 TO Z
SELECT 1
USE BASICDB.DBF
STORE RECCOUNT() TO MAX
GO TOP
IF THEFIELD="N"
Y=100
ELSE
Y=239
ENDIF
REPLACE ALL BASICDB->TEMP WITH SUBSTR(&THEFIELD,2,Y)+SUBSTR
(&NXTFIELD,1,2)
DO WHILE Z<(MAX+1)
STORE 1 TO X
STORE "" TO THESTORE
STORE " " TO THENAME
SELECT 1
USE BASICDB.DBF
GO Z
DO WHILE X<240
STORE " " TO THEAA
STORE " " TO THECODE
SELECT 1
STORE NAME TO THENAME
STORE SUBSTR(TEMP,X,3) TO THECODE
SELECT 2
USE TRANSLAT.DBF
IF SUBSTR(THECODE,1,1)="N".OR.SUBSTR(THECODE,2,1)="N".OR.SUBSTR
(THECODE,3,1)="N".OR.SUBSTR(THECODE,1,1)="O".OR.SUBSTR(THECODE,2,1)
```

```
="O".OR.SSUBSTR(THECODE,3,1)="O"
STORE "X" TO THEAA
ELSE
LOCATE FOR CODE=THECODE
STORE AA TO THEAA
ENDIF
THESTORE=THESTORE+THEAA
X=X+3
ENDDO
REPLACE FSTORE WITH THESTORE
SELECT 3
USE PROT2.DBF
GO Z
REPLACE PROT2->&THEFIELD WITH B->FSTORE
REPLACE NAME WITH THENAME
Z=Z+1
ENDDO
START=START+1
STORE SUBSTR(FLDSTORE,START,1) TO THEFIELD
ENDDO
SELECT 3
USE PROT2.DBF
GO TOP
GO TOP
REPLACE ALL AA1 WITH A+B+C
REPLACE ALL AA2 WITH D+E+F
REPLACE ALL AA3 WITH G+H+I
REPLACE ALL AA4 WITH J+K+L
REPLACE ALL AA5 WITH M+N
USE
```

C.2.3 RF3.PRG

```
*RF3.PRG translates BASICDB.DBF in RF3 and stores data in PROT3.DBF
```

```
STORE 1 TO START
STORE 238 TO Y
STORE "ABCDEFGHIJKLMNA" TO FLDSTORE
STORE SUBSTR(FLDSTORE,START,1) TO THEFIELD
STORE SUBSTR(FLDSTORE,START+1,1) TO NXTFIELD
DO WHILE START<15
STORE 1 TO Z
SELECT 1
USE BASICDB.DBF
STORE RECCOUNT() TO MAX
GO TOP
IF &THEFIELD="N"
Y=99
ELSE
Y=238
ENDIF
REPLACE ALL BASICDB->TEMP WITH SUBSTR(&THEFIELD,3,Y)+SUBSTR
(&NXTFIELD,1,2)
DO WHILE Z<(MAX+1)
STORE 1 TO X
STORE "" TO THESTORE
STORE " " TO THENAME
SELECT 1
USE BASICDB.DBF
GO Z
DO WHILE X<240
STORE " " TO THEAA
STORE " " TO THECODE
SELECT 1
STORE NAME TO THENAME
STORE SUBSTR(TEMP,X,3) TO THECODE
SELECT 2
```

```
USE TRANSLAT.DBF
IF SUBSTR(THECODE,1,1)="N".OR.SUBSTR(THECODE,2,1)="N".OR.SUBSTR
(THECODE,3,1)="N".OR.SUBSTR(THECODE,1,1)="O".OR.SUBSTR(THECODE,2,1)
="O".OR.SSUBSTR(THECODE,3,1)="O" STORE "X" TO THEAA
ELSE
LOCATE FOR CODE=THECODE
STORE AA TO THEAA
ENDIF
THESTORE=THESTORE+THEAA
X=X+3
ENDDO
REPLACE FSTORE WITH THESTORE
SELECT 3
USE PROT3.DBF
GO Z
REPLACE PROT3->&THEFIELD WITH B->FSTORE
REPLACE NAME WITH THENAME
Z=Z+1
ENDDO
START=START+1
STORE SUBSTR(FLDSTORE,START,1) TO THEFIELD
ENDDO
SELECT 3
GO TOP
GO TOP
REPLACE ALL AA1 WITH A+B+C
REPLACE ALL AA2 WITH D+E+F
REPLACE ALL AA3 WITH G+H+I
REPLACE ALL AA4 WITH J+K+L
REPLACE ALL AA5 WITH M+N
USE
```

C.2.4 Structure of the three protein databases PROT1/2/3.DBF

Field	Fieldname	Type	Width	Dec
1	NAME	Character	10	
2	REF	Character	4	
3-16	A-N	Character	80	
17-21	AA1-AA5	Character	240	

C.2.5 FILLAA.PRG

This programme merely takes the protein databases and re-formats and aligns them. It also uses a temporary database TEMPLIST.DBF to store values prior to appending the record to the appropriate database.

```
*FILLAA.PRG
USE AA3MAX
STORE RECCOUNT() TO MUTMAX
USE PROT3
STORE RECCOUNT() TO RECMAX
STORE 1 TO X
STORE "" TO AASTORE
STORE "" TO AAVAR
STORE 0 TO AANO
STORE 0 TO AAFLDNO
DO WHILE X<=MUTMAX
STORE 2 TO REC
USE AA3MAX
GO X
STORE AA_NO TO AANO
IF AA_NO <=240
AAFLDNO=AANO
STORE "AA1" TO THEFIELD
ELSE
IF AA_NO>240.AND.AA_NO<=480
AAFLDNO=AANO-240
```

```
STORE "AA2" TO THEFIELD
ELSE
IF AA_NO>480.AND.AA_NO<=720
AAFLDNO=AANO-480
STORE "AA3" TO THEFIELD
ELSE
IF AA_NO>720.AND.AA_NO<=960
AAFLDNO=AANO-720
STORE "AA4" TO THEFIELD
ELSE
IF AA_NO>960.AND.AA_NO<=1073
AAFLDNO=AANO-960
STORE "AA5" TO THEFIELD
ENDIF
ENDIF
ENDIF
ENDIF
ENDIF
USE PROT3
GO TOP
STORE SUBSTR(&THEFIELD,AAFLDNO,1) TO AAVAR
USE TEMPLIST
GO 1
REPLACE VAR WITH "&AAVAR"
REPLACE AA_NO WITH AANO
DO WHILE (REC<=RECMAX)
USE PROT3
GO REC
STORE REF TO THEREF
STORE SUBSTR(&THEFIELD,AAFLDNO,1) TO AASTORE
USE TEMPLIST
GO 1
IF NOT.(AASTORE=AAVAR)
```

```
REPLACE &THEREF WITH "&AASTORE"
REC=REC+1
ELSE
REPLACE &THEREF WITH "-"
REC=REC+1
ENDIF
ENDDO
USE AA3LIST
APPEND FROM TEMPLIST
X=X+1
ENDDO
```

C.2.6 Structure of the reformatted protein databases AALIST1/2/3.DBF

Field	Fieldname	Type	Width	Dec
1	AA_NO	Numeric	4	
2	VAR	Character	1	
3	A00	Character	1	
3-12	A01-A09	Character	1	
13-14	A_1-A_2	Character	1	
15-19	B00-B04	Character	1	
20	B20	Character	1	
21-28	B05-B12	Character	1	
29-30	C01-C02	Character	1	
31	C00	Character	1	
32-47	C03-C18	Character	1	
48-60	D00-D12	Character	1	
61	D_0	Character	1	
62-67	D13-D18	Character	1	
68-70	E01-E03	Character	1	
71-73	F01-F03	Character	1	

C.3 Determining the consensus sequence of the records in a database

C.3.1 AGCT.PRG

```
*Programme AGCT.PRG, reads sequence data from BASICDB.DBF
*counts the number of times each of the bases A, G, C or T occurs at each position
*converts the count to a percentage of the total records with data at this position
*and writes the results to a database file of filename SEQCT.DBF
STORE 0 TO ARECORD
STORE 14 TO START
STORE "X" TO THEFIELD
STORE 0 TO PLACE
DO WHILE ARECORD<3221
STORE "NMLKJIHGFEDCBA" TO FLDSTORE
DO WHILE (START>=1)
STORE SUBSTR(FLDSTORE,START,1) TO THEFIELD
SELECT 1
USE BASICDB.DBF
GO TOP
STORE 1 TO X
DO WHILE (X <= 240)
SELECT 1
STORE 0 TO THEAS
STORE 0 TO THEGS
STORE 0 TO THECS
STORE 0 TO THETS
STORE 0 TO THENS
STORE 0 TO THEOS
STORE 0 TO TOTAL
STORE 0 TO SPEC
COUNT FOR (SUBSTR(&THEFIELD,X,1)="A") TO THEAS
COUNT FOR (SUBSTR(&THEFIELD,X,1)="G") TO THEGS
COUNT FOR (SUBSTR(&THEFIELD,X,1)="C") TO THECS
COUNT FOR (SUBSTR(&THEFIELD,X,1)="T") TO THETS
COUNT FOR (SUBSTR(&THEFIELD,X,1)="N") TO THENS
```

```
COUNT FOR (SUBSTR(&THEFIELD,X,1)="O") TO THEOS
STORE RECCOUNT() TO TOTAL
SPEC=TOTAL-THENS
SELECT 2
USE SEQCT
PLACE=PLACE+1
IF PLACE<=3221
GO PLACE
REPLACE ACT WITH (THEAS/SPEC)*100
REPLACE GCT WITH (THEGS/SPEC)*100
REPLACE CCT WITH (THECS/SPEC)*100
REPLACE TCT WITH (THETS/SPEC)*100
REPLACE NCT WITH (THENS/SPEC)*100
REPLACE OCT WITH (THEOS/SPEC)*100
REPLACE SPEC_NO WITH SPEC
REPLACE STRINGNT WITH X
ARECORD=ARECORD+1
REPLACE NT_NO WITH ARECORD
REPLACE NT_STRING WITH THEFIELD
SKIP 1
ELSE
X=240
ENDIF
X=X+1
ENDDO
START=START-1
ENDDO
ENDDO
```

C.3.2 Structure of database containing base counts SEQCT.DBF

Field	Fieldname	Type	Width	Dec
1	NT_NO	Numeric	4	
2	NT_STRING	Character	1	

3	STRINGNT	Numeric	4	
4	SPEC_NO	Numeric	4	
5	ACT	Numeric	6	2
6	GCT	Numeric	6	2
7	CCT	Numeric	6	2
8	TCT	Numeric	6	2
9	NCT	Numeric	6	2
10	OCT	Numeric	6	2

C.3.3 AACT.PRG

```
*Programme AACT.PRG, reads sequence data from the protein databases
*counts the number of times each of the 20 amino acids occurs at each position
*converts the count to a percentage of the total records with data at this position
*and writes the results to a database file of filename AACT.DBF
STORE "PROT3.DBF" TO PROTFILE
STORE 0 TO ARECORD
STORE 13 TO START
STORE 12 TO LENGTH
STORE "XXX" TO THEFIELD
STORE "XXX" TO AAFIELD
STORE 0 TO PLACE
STORE "AA1AA2AA3AA4AA5" TO FLDSTORE
DO WHILE ARECORD<1073
DO WHILE (START>=1)
STORE SUBSTR(FLDSTORE,1,3) TO THEFIELD
STORE SUBSTR(FLDSTORE,4,LENGTH) TO FLDSTORE
LENGTH=LENGTH-3
SELECT 1
USE &PROTFILE
GO TOP
STORE 1 TO X
DO WHILE (X<=240)
SELECT 1
```

```
STORE 0 TO F
STORE 0 TO L
STORE 0 TO I
STORE 0 TO M
STORE 0 TO V
STORE 0 TO S
STORE 0 TO P
STORE 0 TO T
STORE 0 TO A
STORE 0 TO Y
STORE 0 TO J
STORE 0 TO H
STORE 0 TO Q
STORE 0 TO N
STORE 0 TO K
STORE 0 TO D
STORE 0 TO E
STORE 0 TO C
STORE 0 TO W
STORE 0 TO R
STORE 0 TO G
STORE 0 TO NOAA
STORE 0 TO BLANK
STORE 0 TO TOTAL
STORE 0 TO SPEC
COUNT FOR (SUBSTR(&THEFIELD,X,1)="F") TO F
COUNT FOR (SUBSTR(&THEFIELD,X,1)="L") TO L
COUNT FOR (SUBSTR(&THEFIELD,X,1)="I") TO I
COUNT FOR (SUBSTR(&THEFIELD,X,1)="M") TO M
COUNT FOR (SUBSTR(&THEFIELD,X,1)="V") TO V
COUNT FOR (SUBSTR(&THEFIELD,X,1)="S") TO S
COUNT FOR (SUBSTR(&THEFIELD,X,1)="P") TO P
COUNT FOR (SUBSTR(&THEFIELD,X,1)="T") TO T
```

```
COUNT FOR (SUBSTR(&THEFIELD,X,1)="A") TO A
COUNT FOR (SUBSTR(&THEFIELD,X,1)="Y") TO Y
COUNT FOR (SUBSTR(&THEFIELD,X,1)="*") TO J
COUNT FOR (SUBSTR(&THEFIELD,X,1)="H") TO H
COUNT FOR (SUBSTR(&THEFIELD,X,1)="Q") TO Q
COUNT FOR (SUBSTR(&THEFIELD,X,1)="N") TO N
COUNT FOR (SUBSTR(&THEFIELD,X,1)="K") TO K
COUNT FOR (SUBSTR(&THEFIELD,X,1)="D") TO D
COUNT FOR (SUBSTR(&THEFIELD,X,1)="E") TO E
COUNT FOR (SUBSTR(&THEFIELD,X,1)="C") TO C
COUNT FOR (SUBSTR(&THEFIELD,X,1)="W") TO W
COUNT FOR (SUBSTR(&THEFIELD,X,1)="R") TO R
COUNT FOR (SUBSTR(&THEFIELD,X,1)="G") TO G
COUNT FOR (SUBSTR(&THEFIELD,X,1)="X") TO NOAA
COUNT FOR (SUBSTR(&THEFIELD,X,1)=" ") TO BLANK
STORE RECCOUNT() TO TOTAL
SPEC=TOTAL-NOAA-BLANK
SELECT 2
USE RF3CT.DBF
PLACE=PLACE+1
IF PLACE<=1073
GO PLACE
REPLACE AAFIELD WITH THEFIELD
REPLACE FCT WITH (F/SPEC)*100
REPLACE LCT WITH (L/SPEC)*100
REPLACE ICT WITH (I/SPEC)*100
REPLACE MCT WITH (M/SPEC)*100
REPLACE VCT WITH (V/SPEC)*100
REPLACE SCT WITH (S/SPEC)*100
REPLACE PCT WITH (P/SPEC)*100
REPLACE TCT WITH (T/SPEC)*100
REPLACE ACT WITH (A/SPEC)*100
REPLACE YCT WITH (Y/SPEC)*100
```

```
REPLACE JCT WITH (J/SPEC)*100
REPLACE HCT WITH (H/SPEC)*100
REPLACE QCT WITH (Q/SPEC)*100
REPLACE NCT WITH (N/SPEC)*100
REPLACE KCT WITH (K/SPEC)*100
REPLACE DCT WITH (D/SPEC)*100
REPLACE ECT WITH (E/SPEC)*100
REPLACE CCT WITH (C/SPEC)*100
REPLACE WCT WITH (W/SPEC)*100
REPLACE RCT WITH (R/SPEC)*100
REPLACE GCT WITH (G/SPEC)*100
REPLACE XCT WITH (NOAA/(TOTAL-BLANK))*100
REPLACE SPEC_NO WITH SPEC
REPLACE AA_NO WITH PLACE
ARECORD=ARECORD+1
SKIP 1
ELSE
X=240
ENDIF
X=X+1
ENDDO
START=START-3
ENDDO
ENDDO
```

C.3.4 Structure of databases containing amino acid counts RF1/2/3CT.DBF`

Field	Fieldname	Type	Width	Dec
1	AA_NO	Numeric	4	
2	SPEC_NO	Numeric	4	
3	AAFIELD	Character	3	
4	FCT	Numeric	6	2
5	LCT	Numeric	6	2
6	ICT	Numeric	6	2

7	MCT	Numeric	6	2
8	VCT	Numeric	6	2
9	SCT	Numeric	6	2
10	PCT	Numeric	6	2
11	TCT	Numeric	6	2
12	ACT	Numeric	6	2
13	YCT	Numeric	6	2
14	JCT	Numeric	6	2
15	HCT	Numeric	6	2
16	QCT	Numeric	6	2
17	NCT	Numeric	6	2
18	KCT	Numeric	6	2
19	DCT	Numeric	6	2
20	ECT	Numeric	6	2
21	CCT	Numeric	6	2
22	WCT	Numeric	6	2
23	RCT	Numeric	6	2
24	GCT	Numeric	6	2
25	XCT	Numeric	6	2

C.3.5 GETMAXNT.PRG

This programme gets the consensus nucleotide at each point along the genome. It uses a temporary database TEMPNT.DBF which has the same stucture as SEQMAX.DBF.

```
*GETMAXNT
STORE 0 TO THEACT
STORE 0 TO THEGCT
STORE 0 TO THECCT
STORE 0 TO THETCT
STORE 0 TO NTNO
STORE " " TO NT_STRING
STORE 1 TO PLACE
USE SEQCT.DBF
STORE RECCOUNT() TO MAX
```

```
DO WHILE PLACE<=MAX
USE SEQCT.DBF
GO PLACE
STORE NT_NO TO NTNO
STORE NT_STRING TO NTSTRING
STORE ACT TO THEACT
STORE GCT TO THEGCT
STORE CCT TO THECCT
STORE TCT TO THETCT
IF (THEACT>THEGCT.AND.THEACT>THECCT.AND.THEACT>THETCT)
STORE "ACT" TO THENAME
STORE &THENAME TO THEVALUE
ELSE
IF (THEGCT>THEACT.AND.THEGCT>THECCT.AND.THEGCT>THETCT)
STORE "GCT" TO THENAME
STORE &THENAME TO THEVALUE
ELSE
IF (THECCT>THEACT.AND.THECCT>THEGCT.AND.THECCT>THETCT)
STORE "CCT" TO THENAME
STORE &THENAME TO THEVALUE
ELSE
IF (THETCT>THEACT.AND.THETCT>THEGCT.AND.THETCT>THECCT)
STORE "TCT" TO THENAME
STORE &THENAME TO THEVALUE
ENDIF
ENDIF
ENDIF
ENDIF
USE TEMP.DBF
REPLACE NT_NO WITH NTNO
REPLACE NT_STRING WITH NTSTRING
REPLACE NT1NAME WITH THENAME
REPLACE NT1 WITH THEVALUE
```

```
USE SEQMAX.DBF
APPEND FROM TEMP.DBF
PLACE=PLACE+1
ENDDO
```

C.3.6 Structure of database containing the nucleotide consensus SEQMAX.DBF

Field	Fieldname	Type	Width	Dec
1	NT_NO	Numeric	4	
2	NT_STRING	Character	1	
3	NT1	Numeric	3	
4	NT1NAME	Character	1	

C.3.7 GETMAXAA.PRG

This programme gets the consensus amino acid at each point along the genome. It uses a temporary database TEMP.DBF which has the same stucture as the three consensus amino acid files AA1/2/3.DBF

```
*GETMAXAA.PRG
STORE 0 TO F
STORE 0 TO L
STORE 0 TO I
STORE 0 TO M
STORE 0 TO V
STORE 0 TO S
STORE 0 TO P
STORE 0 TO T
STORE 0 TO A
STORE 0 TO Y
STORE 0 TO J
STORE 0 TO H
STORE 0 TO Q
STORE 0 TO N
STORE 0 TO K
STORE 0 TO D
```

```
STORE 0 TO E
STORE 0 TO C
STORE 0 TO W
STORE 0 TO R
STORE 0 TO G
STORE 0 TO B
STORE 0 TO AANO
STORE "XXX" TO AA_STRING
STORE 1 TO PLACE
USE RF3CT.DBF
STORE RECCOUNT() TO MAX
DO WHILE PLACE<=MAX
USE RF3CT.DBF
GO PLACE
STORE AA_NO TO AANO
STORE AAFIELD TO AASTRING
STORE FCT TO F
STORE LCT TO L
STORE ICT TO I
STORE MCT TO M
STORE VCT TO V
STORE SCT TO S
STORE PCT TO P
STORE TCT TO T
STORE ACT TO A
STORE YCT TO Y
STORE JCT TO J
STORE HCT TO H
STORE QCT TO Q
STORE NCT TO N
STORE KCT TO K
STORE DCT TO D
STORE ECT TO E
```

```
STORE CCT TO C
STORE WCT TO W
STORE RCT TO R
STORE GCT TO G
STORE XCT TO B
STORE "FLIMVSPTAYJHQNKDECWRGB" TO FLDSTORE
STORE 20 TO XCHECK
STORE 22 TO THELENGTH
STORE 0 TO YCHECK
STORE 0 TO MAXVALUE
STORE 0 TO TEMPVALUE
STORE " " TO MAXNAME
STORE " " TO TEMPNAME
DO WHILE XCHECK>=1
XCHECK=XCHECK-1
YCHECK=THELENGTH-2
STORE 2 TO CONTROL
STORE SUBSTR(FLDSTORE,1,1) TO THECOUNT
DO WHILE YCHECK>=1
STORE SUBSTR(FLDSTORE,CONTROL,1) TO NXTCOUNT
IF &THECOUNT>&NXTCOUNT
STORE THECOUNT TO TEMPNAME
STORE &THECOUNT TO TEMPVALUE
ELSE
STORE NXTCOUNT TO TEMPNAME
STORE &NXTCOUNT TO TEMPVALUE
ENDIF
IF TEMPVALUE>MAXVALUE
STORE TEMPNAME TO MAXNAME
STORE TEMPVALUE TO MAXVALUE
ENDIF
YCHECK=YCHECK-1
CONTROL=CONTROL+1
```

```
ENDDO
THELENGTH=THELENGTH-1
STORE SUBSTR(FLDSTORE,2,THELENGTH) TO FLDSTORE
ENDDO
USE AATEMP.DBF
REPLACE AA_NO WITH AANO
REPLACE AA_STRING WITH AASTRING
REPLACE AANAME WITH MAXNAME
REPLACE AAMAX WITH MAXVALUE
USE AA3MAX.DBF
APPEND FROM AATEMP.DBF
PLACE=PLACE+1
ENDDO
```

C.3.8 Structure of database containing amino acid consensus AA1/2/3.DBF

Field	Fieldname	Type	Width	Dec
1	AA_NO	Numeric	4	
2	AA_STRING	Character	3	
3	AAMAX	Numeric	6	2
4	AANAME	Character	1	

APPENDIX D

DEFINITIVE AND QUASI-SPECIES AMINO ACID DIFFERENCES OVER THE ORFS OF HBV

Table 26. Definitive and quasi-species amino acid differences over the ORFs of HBV

RF1	No	aa	A	A'	B'	P	B"	C	Cq-	D	Dm	E	F	*value 2	*value 3	*value 4
Start preS2	2	S												Y = isolate D10		
	3	T												A = isolate B09		
	4	T	A	A									H	Q = isolate F01		
	5	F												L = isolate A07		
	6	F							C					L = isolate B'04		
	8	A			T	T	T			T	T			T = isolates A02, A03		
	9	L												V = isolate C03		
	10	Q						L	L				L	Q = isolate Cq-02		
	12	P												S = isolate B'04		
	13	P												K = isolates A04, A06	G = isolate B'02	del =isolate D09
	16	G			A	A	A						A	V = isolate Cq-02	del =isolate D09	
	19	F	L											L = isolates B'01, B'04, B05-06, C03, D08	F = isolates A01, A05, A06, A08	del =isolate D09
	28	G												A = isolate Dm14		
	29	V		L									Q	L = isolate E03		
	30	N			S	S										
	32	V	A		A	A							A			
	33	F			Q	Q	Q							L = isolates C15, D13, Dm17, Dm18		
	34	T	N	N	N	N	N							N = isolate E03	T = isolate A'02	
	35	T	I	I									I	I = isolates C14-C17		
	36	A			V	V	V							V = isolates C11, C14-15, D08-D12	A = isolate B"11	
	37	S												Y = isolate B'03		
	38	F	H	H	A	A	S		L			L	L	H = isolates D09, Dm15, E03	T = isolates B05-06, D06-07	
	39	I									I		T			
	42	I												V = isolate E03		
	43	F	S	S	L	L	L						S	S = isolates B'04, C07, Dm17, Dm18	F = isolate F01	
	44	S	A											A = isolates Dm17, Dm18	T = isolate D09	
	45	F			K	K	K						K	T = isolates B"11, E03		
	46	T								I	I	I		I = isolates A'02, Cq-02	T = isolates Dm17, Dm18, E03	
	48	D											G			
	50	A	V		V	V	V		V					V = isolates D03, Dm16, Dm18		
	51	P	T	T					T	I	L		M	T= isolates B08, Dm16, Dm18	P = isolate D10, A'02	
	52	N												I = isolates A07, Dm18		
Start S	54	E											D	E = isolate F02		
	55	N						S				S		S = isolate B'01	N = isolates C03, C05-C09, C18	T = isolate C04
	56	I						T	T							
	57	T			A	A	A							T = isolate B'03		

Table 26 continued overleaf

RF1	No	aa	A	A'	B'	B	B*	C	Cq-	D	Dm	E	F	value 2	value 3	value 4
	59	G												D = isolate D05		
	60	F			L	L	L						L	F = isolates B05-06		
	63	F												H = isolate B04		
	65	L												P = isolate A01	R = isolate F02	
	67	L												S = isolate D05		
	68	Q												E = isolate Dm18		
	70	G											V	V = isolates C16, C17		
	71	F											C			
	72	F												L = isolate A05	S = isolate A09	
	73	L				S	S							S = isolates B04, C05, C16-17	L = isolates B07 B10, B"12	
	76	F			K	K	K					K	K	K = isolates Cq-02, D08		
	78	L												H = isolates C16-17		
	79	T												A = isolate C15		
	81	F												L = isolates C16-17		
	82	Q												L = isolate B"11	K = isolates D08, D10	
	85	D												H = isolate C08		
	87	W												G = isolate Cq-01		
	88	W												del = isolate C18		
	92	N												S = isolate B08		
	94	L												I = isolate A05	P = isolate B06	R = isolate Dm16
	96	G												E = isolates B"11, Cq-01, C13-15	A = isolate C08	
	97	T	S	S				A	A			A	I		A = isolate A'02	
	98	F								T	T					
	99	V						T					G	R= isolates Cq-01, F02	A = isolate B04	
	101	L						F	F				P	P = isolate B04	L= isolate C08	
	103	Q												P = isolate C03		
	106	Q												R = isolate A01		
	108	F			Q	C	Q									
	109	T			I	I	I									
	111	N			S	S	S							S= isolates D03, E02	H = isolate Dm18	
	113	S											L			
	116	S			C	C	C									
	120	I								T	T		T	T= isolates B'01, C05, C10	I = isolate Dm18	
	131	R												G = isolate Dm16		
	132	F												S = isolates C16-17		
	137	F			C	C	C									
	148	V												A = isolate C18		
	150	L												V = isolate C05		
	151	D												E = isolate Dm16		
	153	Q												K = isolates C16-17		
	157	F												A = isolates B05-06	S = isolate C08	H = isolate Dm16
	161	L												I = isolate A09		
	162	I						L	L				L	I = isolate F03	T = isolate B"04	
	163	F												L = isolate A01		
	164	G												R = isolate C08		
	165	S						T	T					T = isolates Dm16, Dm18		
	166	S	T	T									T	T = isolates B"04, Dm16		
	170	T												A = isolates D02, D04-05	V = isolate D01	
	174	K								R	R	R		K = isolates Dm16, E03	R = isolate C10	
	177	T												M = isolates D10, Dm13, Dm14, E01		
	178	T						I	I					A = isolates B"01, F02		
	179	F										I	L	T = isolates D01-02, D04-D05, Dm13-Dm14	P = isolate E03	
	180	A												V = isolate D01		
	181	Q												R = isolate B06		
	182	G												N = isolate B06		
	183	T	N	N										N = isolates D09, Dm16		
	185	M												L = isolates E03, A09	K = isolate A07	
	186	F								Y	Y			F = isolate Dm16	L = isolate C05	I = isolate E03
	189	C												S = isolate C08		
	190	C												W = isolate C12		
	192	T										S	S	T = isolate E03		
	194	F												T = isolate Dm16		
	195	S	T	T	T	T	T							M = isolate B10		

Table 26 continued overleaf

Rf1	No	aa	A	A	B'	B	B*	C	Cq-	D	Dm	E	F	value 2	value 3	value 4
	196	C												V = isolate A05		
	210	F											L			
	211	A							V	G	G	G	G	A = isolates Dm16. E03		
	212	K						F	F							
	213	F	Y	Y	Y	Y	Y						Y	Y = isolate Dm16		
	216	E												G = isolate B'04	Q = isolate Dm18	
	217	W												R = isolate A09	G = isolate C07	
	220	V								A	A	A	A	V = isolates Dm16. E03		
	229	V							A					A = isolate E03		
	230	F											Q			
	235	F											C			
	236	V						A				A		V= isolates C03-C10		
	244	L												V = isolate Dm18		
	245	S											L	L = isolates A09. E03, D08		
	246	V	A	A										A = isolates C10, Dm16, Dm18, E03	L = isolate A05	
	250	M											I	I = isolate B'02		
	252	Y				F	F									
	254	G												E = isolate F03		
	255	F												Q = isolates C16.17		
	256	S											N	N = isolates E03, D06, D07	G = isolate C17	K = isolate D08
	258	Y											C	F = isolate D08		
	259	N	S							S	S		S	T = isolate D08	D = isolate E01	
	260	I												T = isolates B05,06		
	261	L	V											V= isolate Dm18		
	262	S												R = isolates A09, B'04		
	263	F												S= isolate A01		
	265	I			M	M	M	L	L	L	L			I = isolates B'01, C18, Dm18		
	270	I												T = isolate B'04		
	272	F											C	L = isolate C18		
	273	C											Y	Y = isolates C05, 06		
	276	V												A = isolates B06, D09		
ans S	277	Y											S	F = isolate C17		
Start	610	H												P = isolate C10		
pneC	614	I				V								V = isolate E03		
	616	I												F = isolates D12, F01		
	617	C												R = isolates B05, B06		
	618	S	T	T										S = isolate A09	L = isolate F03	T = isolate Dm16
	620	F												L = isolate A09		
	622	V		F										F = isolate C17		
	632	L												P = isolate C15		
	633	W				*				*				STOP = isolates A09. B'04, C04. C05, C16-18, D17, E03	W = isolates B07, B10. D01-03. D05, D11 D12	R = isolate C03
	634	G												D = isolates B'01, B05, B06, C17, C18,		
Start C	639	F												T = isolate C03		
	642	E												Q = isolate Dm17		
	644	G												E = isolate A'02		
	645	A												L = isolate A'02		
	646	S	T	T		T				T	T	T		S = isolates A09, B05, B06, B08. D06, D07, D09	L = isolate A'02	
	647	V												W = isolate A'02		
	648	F												S = isolate A'02		
	649	L												Y = isolate A'02		
	650	L											F	V = isolates C16, C17	L = isolate F01	S = isolate A'02
	651	S												H = isolate A'02	P = isolate Dm17	
	652	F												Y = isolate Dm17		
	653	L												C = isolate A'02		

Table 26 continued overleaf

No	aa	A	A	B'	B	B*	C	Cg-	D	Dm	E	F	value 2	value 3	value 4	value 5
654	P												L = isolate A'02	A = isolate F02		
655	S												A = isolates B05-06	L = isolate A'02	H = isolate A09	T = isolate D06
656	D												T = isolate A'02			
657	F												S = isolate A'02			
658	F												Y = isolate D09			
659	P												L = isolates A'02, Dm17			
660	S												P = isolate A'02	Y = isolate Dm17	A = isolate B'01	
661	V			I		I	I	I					V = isolates B'01, B'02	I = isolate F03	L = isolate D03	A = isolate D01
662	R												H = isolates B05, B06	G = isolate A'02		
663	D												I = isolate A'02			
664	L												Y = isolate A'02			
665	L												STOP = isolate A'02			
666	D												N = isolate A'02			
667	T												N = isolate D11			
668	A												V = isolate B10			
669	S				A								S = isolates B05, B06, B10, C16, C17	T = isolates D06, D07	F = isolate C03	
672	Y												F = isolate D10	H = isolate C03		
674	E								D		D	D	E = isolates D08-12, E03	D = isolate Dm15		
677	E												K = isolate B10			
680	E												Q = isolate B08			
683	S											F	T = isolates C16, D02, D06, D07, D09			
684	P												A = isolates C16, C17			
685	H											N	N = isolate E03			
689	L												I = isolates B'01, B05, B06			
692	A												Q = isolate B07			
693	I							V					V = isolate A09			
694	L												V = isolate C03			
697	G												V = isolate A04			
698	E												D = isolates D04, D08, D10	K = isolate A'02		
699	L												V = isolate D10			
700	M												I = isolate D10			
701	T			N		N	N	N					N = isolates D04, D10			
703	A												S = isolate D08			
704	T											S	S = isolates B05, B06, E03			
706	S	N	N		N				G	V	V	N	V = isolates D01-05	G = isolate F04	N = isolate E03	D = isolate Dm18
711	E				Q								Q = isolates A07, A09, B03, D03, D04, D06, D07, D09	E = isolates B07, B09, B10		
713	P															
714	A												T = isolates C18, D08-10,	I = isolate D11		
715	S											A				
717	D			E		E	E	E					E = isolates B05, B06, D04, E03	D = isolate B'01		
718	L												Q = isolates A09, E03	A = isolate C18		
719	V												I = isolate D10			
721	S	N	N		N							N	N = isolates D04, D10, E03	G = isolate C04		
725	T			V		V	V	V					N = isolate A06	A = isolate B03		
726	N												H = isolates B05, B06			
727	M												V = isolates A02, A03			
730	K												R = isolate D10			
731	I								F	F	F		L = isolates B04, C04, C05, C16-C18	I = isolates Dm18, E03	F = isolate B01	
734	L												I = isolates C05, C16, C17			
735	L												W = isolate B10			
739	I												V = isolates C04, D04, D09	L = isolate D10		

Table 26 continued overleaf

No	aa	A	A'	B'	B	B"	C	Cq-	D	Dm	E	F	value 2	value 3	value 4	value 5
742	L												I = isolates B05, B06			
743	T												M = isolates A09, C16, C17			
747	P												D = isolates A09, C16	Q = isolate B'01		
750	L									I	I		L = isolates Dm18, E03	V = isolate D09		
755	S												A = isolate Dm16			
756	P												V = isolate C03			
761	P												G = isolate C14			
764	P												T = isolates B'04, D05, Dm16-18	Q = isolates C03, D04, D09, D10		
769	P												Q = isolate B09	S = isolate D08	T = isolate D09	
780	T										N		N = isolate B'04	T = isolate E03		
781	T												C = isolates D06, D07, D09	A = isolate B'04		
783	V												I = isolates C05, C16, C17, D04, D06, D07			
784	P												T = isolate Dm17			
785	P												Q = isolates B'01, C16, C17,D03, D06, D07	C = isolate A09		
787	-	D	D													
788	-	P	P													
789	G												R = isolates A09, C16, C17	C = isolate B09	D = isolate Dm18	S = isolate C15
791	S												T = isolates B'01, D04, D06, D07.			
800	P												A = isolate Dm17			
802	P												P = isolate Dm17			
804	S												P = isolate D10			
805	Q												P = isolates A02, A03	K = isolate C03		
808	P												G = isolate B"12			
812	S												T = isolate D08			
813	Q												R = isolates B05, B06			
815	P										P	P	L = isolate D03			
816	E										A	A	G = isolates B'04, D10			
817	S				P								P = isolates C18, Dm14	S = isolates B07, B'09.		
459	M												I = isolate A'02			
460	A												G = isolate A08			
461	A												T = isolate A01			
463	L						V	V				M	V = isolate B'04	L = isolates C10, F02	M = isolates C17 C18	
464	C	Y	Y										Y = isolates Dm16-18	C = isolates A01, A09		
466	Q												K = isolate F01			
469	P												S = isolate A'02	T = isolates B08, E03		
470	A	S	S					T					S = isolates Dm17-18, E03	E = isolate D09		
472	D												A = isolate A'01			
475	C												R = isolate D09			
478	P												S = isolate B'12			
479	V												I = isolate F01			
480	G										S		G = isolate E03			
484	R										E	S	C = isolates D05, D10, D13-14	R = isolate F01	S = isolate D09	H = isolate D12
485	G												R = isolate C13			
487	P											T	P = isolate F01			
488	L						V	V	F	F	V		V = isolates D06-07, D15	F = isolates B'01, E03	I = isolates C04, C17	
[illegible]	G			F	F	F						P	P = isolate C17	A = isolate A'02		
490	G												R = isolates A'02, C02			
491	P									S	S		S = isolates C16-18, D10-15, F02	P = isolate E03		
492	I						F	F								
49[illegible]	G												R = isolates A01, B08, C04, E03			

Table 26 continued overleaf

No	aa	A	A	B'	B'	B"	C	Cq-	D	Dm	E	F	value 2	value 3	value 4	value 5	value 6
494	T			A	A	A	F		D		D	A	T = isolates B'04, C05, C06, C07, C16, C17, D01, D08-11	A = isolates Cq-02, C09, D12, E03	D = isolate Cq-01		
495	L											V	V = isolate A'02	L = isolate F03			
496	P	S	S						S	S	S		S = isolate C08	P = isolate E03			
497	S			F	F	F						P	P = isolate E03				
498	F			A	A	A						S	S = isolate E03	P = isolate F02	L = isolate C11		
499	S												A = isolate C10	H = isolate C03			
500	P						S	S					L = isolate Cq-02, Dm16	F = isolate C11	S = isolate F02	A = isolate F03	
501	S			F		F							S = isolate B12				
502	A			V									A = isolates B03 04	L = isolate B03	T = isolates C08, D12		
503	V												D = isolate A'01	L = isolate E03			
504	F								S				S = isolate Dm16	P = isolate D02			
505	A			T	T	T			T	T			S = isolates A'02, B08, B10-11, D02-05, D09	A = isolates Dm17-18	T = isolate E03		
506	D							N					N = isolates C05, C06				
507	H												Y = isolate E03	D = isolate F01	N = isolate F02		
508	G												R = isolates A'01, C13				
509	A											S	T = isolate A'01				
512	S												Y = isolate A'01				
514	F												P = isolate C13				
515	G												A = isolate C13				
517	F												F = isolate C07	T = isolate Dm18			
518	V												D = isolate A'01				
520	A											S	S = isolate E02				
525	G												V = isolate C13	E = isolates C14, C15			
526	F												A = isolate A'01				
530	P												P = isolate C08				
531	F												L = isolates C13-C15				
535	R												H = isolate Dm17				
536	P	C	C										C = isolates Dm16, E03	S = isolate C10			
538	P												A = isolates A06, A07				
540	T												S = isolates C05, C06				
542	N												K = isolate D11				
543	A												G = isolates B09, C11				
544	H											P	P = isolate E03	Q = isolate D10			
545	Q			F	F	F		M				W	R = isolates D14, E03	G = isolates B'02, B'04	H = isolates D02-D05	L = isolate D10	P = isolate D11
546	V			N	N	N	V	V	F			S	M = isolates D06, D07, D14	V = isolate A'02	S = isolates A09, D18	N = isolate E03	I = isolate D05
547	L												M = isolate Dm17				
549	K											T					
550	V												L = isolates C11, C12				
552	H												Y = isolates C04-06, C18, D10				
554	P												S = isolate C08				
555	T												A = isolate D05				
559	S	P	P										P = isolates D05, Dm16-Dm18	S = isolates A04, A'02			
560	A											G	V = isolates D01, D04, D08, D10, D11				
561	M											R	W = isolate F01				
563	T											M	P = isolate B"12				
565	D											E	D = isolate F02	T = isolate E03			
569	Y												H = isolate B"11				
570	F											I					
572	D												G = isolate A'01				
573	C												S = isolate B'03				

Table 26 continued overleaf

No	aa	A	A	B'	B	B*	C	Cq-	D	Dm	E	F	value 2	value 3	value 4	value 5
574	V						L		L	L	L		V = isolates D05-07, Dm16-Dm18, C09, E03			
576	K			N	N	T							T = isolates B'01, B05, B06, B09			
577	D			E	E	E							E = isolate D09	G = isolate A'01		
581	I												F = isolate D04	S = isolates D08, E02		
583	F												Q = isolate A'01			
584	E												D = isolate D10	del = isolate C06		
585	L												V = isolates B'01-02, Cq-02, D01	N = isolate D10	L = isolate C18	del = isolate C06
586	F												del = isolate C06			
587	I												del = isolate C06			
588	K												M = isolates A06-07, A09, B'01, C03-05, C18, D08, D10		del = isolate C06	
589	V												I = isolates A06-07, A09, B'01, C04-05, C16-18, D08, D10	L = isolate D09	del = isolates C06, B07	
590	F												C = isolate A08	Y = isolate D10	del = isolates C06, B07, C13-C15	
591	V												L = isolate B07		del = isolates C06, C13-C15	
592	L												del = isolates C13-C15			
598	K												del = isolate C09			
599	L												del = isolate C09			
600	V												del = isolate C09			
601	C												del = isolate C09			
602	A			S	S	S	S					S	T = isolate E03	del = isolate C09	A = isolate F03	
603	P												del = isolate C09			
604	A												V = isolates C16, C17	L = isolate A'02	del = isolate C09	
605	P												S = isolate A'02	del = isolate C09		
606	C												del = isolate C09			
610	T												P = isolate C10			
775	Q											P				
776	H												Y = isolate A'01			
778	R												Q = isolate A'01			
779	K								R	R	R		K = isolates D02, D04-05, D14, Dm18, E03	I = isolate A'01		
780	I										I		P = isolate A'01	M = isolate B*04	L = isolate E03	
781	L												V = isolates D06, D07, D09			
784	D												H = isolate Dm17			
785	D			P									V = A'02, Cq-1	K = isolate B'01	N = isolates C17, C18, D03, D06, D07	
786		G	G										E = isolate A'02			
787		T	T													
788	E												D = isolates A09, B09, C16-C17	G = isolate Dm18		
789	A												T = isolate Dm18			
798	P												H = isolate B*12			
800	A												P = isolate Dm17			
802	P	A	A										Q = isolate Dm17			
803	G	D	D						D		D		G = isolates D06, D08-12, E03	D = isolate F01		
805	N												H = isolates A02, A03			
806	P				H								H = isolates Cq-01, D06-07			
807	P												P = isolate B*12	H = isolate D08		
810	E												A = isolate F01			
811	D												E = isolates A'01, D08			
813	N												G = isolates B05, B06			
815	G										Q	Q	W = isolate D03			
816	N										L	L	D = isolates B'04, D10			
817	L				F			F			F	P	L = isolates B05, B06	P = isolate A'01		
820	S												N = isolate C18			
821	I												T = isolate C04	V = isolate C18		

Table 26 continued overleaf

No	aa	A	A	B'	B	B*	C	Cq-	D	Dm	E	F	value 2	value 3	value 4	value 5	value 6
829	N												D = isolate A'01				
830	F												L = isolate A'01				
833	I												F = isolate Dm16				
838	V										I		A = isolate A01	L = isolate E03			
840	V	I	I	C		S						A	I = isolates B08, C06	V = isolate A08			
841	F												L = isolate C11				
844	F								H	H	N	N	N = isolates B01, D10	K = isolates B02, B*12	Q = isolates B*04, B*11	D = isolates C10, C16-17	H = isolate F03
845	W												S = isolate C11	C = isolate C12			
846	Q								K	K	K	L	E = isolate Dm16	K = isolate C09	Q = isolate E03	S = isolate F01	
847	T												S = isolate A08	P = isolate C14			
852	N	K	K	D	D	D	H	D			D	D	N = isolates B05, B06, C03, C11-13, D15-18	S = isolate B04			
856	Q								H	H	H	H	H = isolate A'02	R = isolate D01			
857	E								Q	Q	Q	Q	P = isolate D05				
859	I											L	L = isolates B05, B06	M = isolate F02			
860	I			V	V								I = isolates B05-06	A = isolate A'02			
861	I			D	D				K	K		S	D = isolates A05, A'01, C16-C18	N = isolates D03, D06-07			
862	F								K	K	K	K	K = isolate B*11				
864	Q			K	K	K			E	E	E	E	E = isolates B*12, C06	R = isolate A'01			
866	F						Y	Y									
872	V											K					
875	K			N	N	N						L	K = isolate B*11				
879	K								Q	Q	N		K = isolate E03				
881	I										V	V	S = isolate E03				
884	A												V = isolate D09				
887	Y										F		F = isolate F01	Y = isolate E03			
889	N	T	T							K	I	K	K = isolates D01, D05, D08	T = isolate B09	M = isolate D09	N = isolate A'02	
890	V	H	H				L	L			S		F = isolates C16-17	S = isolate A'02	H = isolate F02	L = isolate F03	
894	L											F	F = isolates A08, C16-17				
895	F												L = isolate F02				
896	I												M = isolate F01				
897	D										E	E	E = isolates B05-06, C05				
899	G												D = isolate F03				
902	F												T = isolate A'02				
906	E	D	D								D		D = isolate F02				
907	H	Q	Q	Y							D		Q = isolates Cq-02, F02	Y = isolates C10, D11	H = isolate A'02	N = isolate D10	
908	V						A	A	L	L		A	T = isolate Cq-02	I = isolate B04			
910	N												D = isolate B*04				
914	Q						K	K					K = isolate F01	Q = isolate C18			
915	T												A = isolates B05, B06, D10				
924	K												Q = isolate E03				
927	I												S = isolate Cq-2	V = isolates D08, D10	H = isolate E01		
930	K												R = isolate A'01				
932	F												G = isolate A'01				
933	T				S	S						S	T = isolate F02	P = isolate A09			
935	F				H				H	H			H = isolates A01, C16-17	R = isolates B05, B06, D02-D07			
940	C												S = isolate A09	F = isolate D10			
943	P												Q = isolates D05, F02				
946	W												S = isolate A'01				
948	Q												H = isolate D10				
44	E												D = isolates B*11, E03				
48	R												R = isolate A01				
50	A	D	D	T			N	T			T	Y	T = isolates A'02, D02, D09, Dm17	A = isolates B*01, C18,	D = isolates B08 Dm16, Dm18	P = isolate D10	
51	F												D = isolates A07, Dm18				
52	H												Y = isolate Dm16				

Table 26 continued overleaf

No	aa	A	A'	B'	B	B"	C	Cg-	D	Dm	E	F	value 2	value 3	value 4	value 5	value 6 & 7
54	G										K	Q	D = isolate C04	E = isolates E03, F02			
56	H			R	R	R	N	N					N = isolates D06, D07	Y = isolate F01	H = isolate F05		
58	I			T	T	T						T	I = isolates B05, B06				
64	A												S = isolates Dm13-15				
67	T												R = isolate Dm18				
75	K												C = isolates B05, B06				
78	H												R = isolate C15				
81	A			T	T	T	T	T				T	S = isolate B11	E = isolates D06, D10			
84	R												T = isolate C08				
86	V												G = isolate Cg-1				
96	N	I	I				S	S			S	T	S = isolate A102	V = isolate A04, A05	H = isolates B05, B06		
97	T								H	V	S		Y = isolates D01, D03, D09-12	S = isolate A03, A04	T = isolate E06		
98	R						H						K = isolate Cg-1	C = isolate F02			
100	S												P = isolates A06, E06				
110	Q												R = isolate Dm18				
118	N					D											
121	S												T = isolate C06				
130	A												G = isolate Dm16				
134	I			L	L	L	L		L	L	L	L	I = isolates D04-07, Dm18				
140	A												S = isolate D02				
146	V	I	I										V = isolates A05, A02				
148	S												C = isolate C05	T = isolate D02			
151	L												V = isolate Dm16				
152	S						P	P					P = isolate F02	C = isolates C16, C17			
156	A												G = isolates B05, B06	V = isolate C06			
157	R												C = isolate C06				
158	L												V = isolate D02				
160	S												P = isolate B05				
161	N						T	T				T	N = isolate F02	H = isolate B04	D = isolate D06		
163	R												K = isolate C06				
164	I						N	N					N = isolate Dm16	Y = isolate Dm18			
165	I	N	N						F	L		H	N = isolates B04, Dm16	F = isolates Dm17, Dm18	H = isolate A101	Y = isolate F03	L = isolates D04-05, E03 I = isolate D09
166	N											D	D = isolate D10				
167	H	N	N	N			V	V		N			H = isolates B01, C08, D14-15	Y = isolates D10-12	N = isolate D01		
168	H	Y	Y										Y = isolates Dm16, E02	R = isolates D01-02, D04-D05	H = isolate A102		
170	G			R		R							R = isolate C05	G = isolates B01, B04			
171	T												A = isolates B08, C09, C15				
172	M		L										D = isolates B01, D03, D06-07, D09, D13	N = isolate C16			
177	D			N	N												
198	R	W	W		W								W = isolates B02, C04, Dm16	R = isolates A101, B08, B09			
205	I												V = isolate C18				
208	I	V											V = isolate Dm16				
207	L										M						
215	G												A = isolate D18				
218	V												A = isolate A09	G = isolate C07			
219	S												G = isolate C15				
242	L												V = isolates C16, C17				
260	A												G = isolate Dm18				
265	S												T = isolate D06				
250	V											L	M = isolate B02				

Table 26 continued overleaf

RT3	No	aa	A	A'	B'	B	B"	C	Cq-	D	Dm	E	F	value 2	value 3	value 4	value 5
	255	K												R = isolate C17			
	256	S												T = isolate D08			
	258	Q												H = isolate D08	R = isolate E01		
	260	L	F											R = isolate Dm18			
	262	S												A = isolates A09, B04	F = isolate A02		
	264	Y						F		F	F			Y = isolates C18, Dm18			
	265	T			A	A	A							T = isolate B'01			
	266	A						S	S					S = isolate E01			
	267	V						I	I					V = isolate C04			
	269	N												H = isolates B06, C05, C06			
	272	L												V = isolate C18			
	274	L											V				
	280	P											T				
	281	N			H	H	Q						S	H = isolate C03	N = isolate B10	A = isolate F02	
	289	S											N	S = isolate F01			
	291	N								H			H	H = isolate Dm16	N = isolates D01, D03-D05		
	294	G												V = isolate A04			
	296	V	I											I = isolates D06-07, Dm17-18	R = isolate A09	V = isolates A01, A08	
	299	S								C	C			C = isolates C05-08	S = isolates D02, D10, D14 Dm17, Dm18		
	300	W								Y	Y			W = isolates Dm17-18)	H = isolates D06-D07		
	302	T								S	S	S		S = isolates B05-06, F01	T = isolates Dm17, Dm18	A = isolate F03	
	303	L												W = isolate B10			
	306	E	D	D						D			D	D = isolates B10, C04, Dm17-18			
	309	V								I	I	I		R = isolates D06-07, D09	V = isolates Dm17-18, E03		
	310	Q					L	L	L				H	Q = isolates C10, C16-18	R = isolate B'03	P = isolates B05, B06	
	311	K												N = isolate B10			
	312	I						L						K = isolates C03-C08, C18	L = isolate B*12	F = isolate B10	
	314	Q	H	H	M	L				F	E		E	H = isolates Dm18, E01	D = isolates E02, F02	N = isolate E03	T = isolate B01
	315	C												W = isolate B'02			
	317	F												G = isolate E01			
	319	L											V	V = isolate Dm16	L = isolate F01		
	321	V												I = isolates A'01, D02, D13, D14			
	323	F												S = isolate C08			
	329	V												G = isolate D09			
	330	C												W = isolates Dm17, Dm18			
	331	Q												P = isolates B05, B06			
	332	R												G = isolate E01	S = isolate F02		
	333	I												L = isolates A'02, C07			
	352	M												K = isolates C05, C06, C16, C17, D01, D06, D07, D10			
	356	A												T = isolate E01	N = isolate F03		
	358	I												T = isolate E02			
	359	Q											T				
	360	A						S		S	S	S		A = isolates D03-05, Dm17, Dm18, C05-06, C18, E03	F = isolate Dm16		
	363	A												T = isolate D03			
	365	T											V	S = isolate A09			
	366	F												L = isolate C15			
	368	F												S = isolates Cq-1, C18			
	372	A												V = isolates D03-05			
	375	C	S	S	S	S								S = isolates Dm17-18, E03	T = isolates B'02, B*12	N = isolate B*11	
	376	K												Q = isolates C11-15, E03	N = isolate C04		
	379	L	M	M									M	M = isolates B03, B10, D06, D07, Dm17-18	L = isolates A09, A'01	S = isolate E03	
	380	N				T								T = isolates D06, D07, E03	H = isolates C03, C11-C15, C18, D10	N = isolate B05	
RT3	386	R												G = isolate D15			

Table 26 continued overleaf

RT3		aa	A	A'	B	B	B*	C	Cq-	D	Dm	E	F	value 2	value 3	value 4	value 5	value 6	value 7
Start	389	F						S	S					S = isolate E03					
RNAse H	396	A												R = isolates B'03, B'04	G = isolate C12				
	406	A								V				V = isolates Dm13-14, E03	G = isolate Dm16	P = isolate B'03			
	407	I								M	M			M = isolates E03, Cq-02	I = isolates Dm17, Dm18				
	409	H										I		H = isolate E03	N = isolates Cq-01, Dm17	Q = isolate C03			
	410	Q						R						Q = isolates C16-17	S = isolate C03				
	411	F												G = isolate C03					
	412	M												V = isolates D10, D14					
	414	G												E = isolate A'02					
	417	V								L	S			P = isolates D06, D07	E = isolate Dm18	Q = isolate D08	M = isolate A'01	R = isolate D09	V = isolate Dm17
	418	A			S	S	S							A = isolateB'03	V = isolate C11				
	419	F												R = isolate D11					
	425	A												V = isolate B'11					
	435	S												R = isolate C07	D = isolate C11				
	436	F												W = isolate C05					
	440	K								N	N		I	N = isolates A'01, B10, Cq-02	K = isolate Dm17	T = isolate F01			
	441	I								I	I			I = isolate Cq-02	L = isolate Dm17				
	442	I								L				L = isolate Dm16	I = isolate D10				
	445	D												H = isolate B'04					
	447	S												F = isolate D09					
	451	S												T = isolate A01					
	452	F												Q = isolate A'02					
	457	F												Y = isolates D11, D14					
	460	L												V = isolate A08					
	464	A	T	T										T = isolates Dm17, Dm18	A = isolates A01, A09				
	467	W												STOP = isolate A'02					
	470	F												S = isolate D09					
	471	G												R = isolate A02					
	484	A												T = isolate D12					
	486	D											E	D = isolate F01					
	494	L								I		I		I = isolate Cq-1	V = isolates A04, A09	L = isolates D01, D08-12			
	495	Y	S							F	S			S = isolates D01, D08-12	F = isolate E01	Y = isolates A05, A09			
	499	L												I = isolate E03					
	500	R						H						R = isolates C04, C16-17	C = isolates Cq-02, Dm16				
	502	P			L				R					R = isolate E01	P = isolate B'04	L = isolate E03			
	503	F		V	Y	Y	Y												
	504	R	Q									Q	Q	Q = isolates B'04, C11-12, Dm17-18	R = isolate E01				
	510	T												A = isolate C03					
	514	A												P = isolate C13					
	515	V			D	D	D			D	D		D	V = isolates Dm17-Dm18	L = isolate C13				
	516	S												F = isolate C07					
	517	P												L = isolate C07					
	522	H				R	R							H = isolate B10					
	525	D	V	V										V = isolates C03, C07, C12, D02, Dm17-18, E03, F03	G = isolates C05, C06	S = isolates C14, C15	Y = isolate C13		
	526	R												L = isolate D15					
	530	A												P = isolate C08					
End	535	H												I = isolate Dm17					
RNAse H	538	R												K = isolates Cq-2, C05, C06					
Start pol 5	953	G								del	del						del = isolate E03		
	954	G								del	del	I	A				del = isolate E03		
	955	W								del	del	S	P	G = isolate A09	P = isolates B05, B06	Y = isolate C05	del = isolate E03		

Table 26 continued overleaf

RF1	No	aa	A	A	B'	R	B*	C	Cq	D	D	E	F	value 2	value 3	value 4	value 5	value 6
	956	S								del	del	W	L		del = isolate E03			
	957	S								del	del	T			del = isolate E03			
	958	K								del	del	V	T		del = isolate E03			
	959	F								del	del				del = isolate E03			
	960	F								del	del			P= isolate F01	del = isolate E03			
	961	K					Q	Q		del	del	F	R	del = isolate E03				
	962	G								del	del			del = isolate E03				
	963	M								del	del	W		del = isolate E03				
	965	T								Q	Q	K	Q					
	966	N												S = isolate D10				
	967	L										I		L = isolate E03				
	968	S												A = isolate B'01				
	969	V								T	T	T						
	970	P								S	S	T		S = isolate E03				
	976	F												L = isolates A07, F02				
	978	C												H = isolate A02	E = isolate E03			
	984	A											L					
	985	F												S = isolate Dm16				
	986	G	G	G	K	K	K	G	G					K = isolate E03				
	989	S								T	T	T						
	990	N			E	E	D			A	A	F	S	D = isolate B'01	H = isolate C07	N = isolate E03		
	991	N											S					
	996	F			L	L	L					H		Y = isolate D01	F = isolate E03			
	998	F											T	T = isolate B'01	K = isolate F03			
	999	N	I	I	H	H	H							N = isolate A03-04, B'02	K = isolate E03			
	1002	H	H	H				H	H	T	T	H	S	D = isolates A02, A03	T = isolate F03	Y = isolates B'01, E03	Q = isolates Cq-01, C09	S = isolate D08
	1004	F											T	P = isolate E03				
	1005	D	A	Q				F	E			E	M	D = isolate C12				
	1006	A					S							G = isolate C12				
	1007	N				H								W = isolate B'01	I = isolates C11-12	T = isolate E03		
	1008	K	Q	Q				Q	Q					K = isolates C11-12				
	1009	V												D = isolate Dm16				
	1011	A	V	V	V	V	V					V	V	V = isolate C18				
	1013	A											G	D = isolates C11-12				
	1014	F											Y					
	1016	F								L	L							
	1017	G												R = isolates A02-A04				
	1018	F	L											L = isolates D09-10, Dm16	F = isolate A01			
	1025	L	I	V										V = isolate A02				
	1028	W												S = isolate A02	L = isolate E03			
	1031	Q												H = isolate B'03				
	1035	I			L							M	V	M = isolates A01, B07, D06-07	A = isolate C04	V = isolate C07	T = isolates C16, C17	I = isolate E03
	1036	L												I = isolates D01-05				
	1037	T		A						Q	Q	K		H = isolates Dm16, Dm17	T = isolate E03	E = isolate D11		
	1038	T				S								T = isolates B05-06				
	1039	V								L	L	L	L	L = isolates Cq-1, C10	V = isolates D09, Dm16, Dm17			
	1040	F	S															
	1041	A	T				T							V = isolates C04-08				
	1042	A	I	I						N	N	D	D	M = isolate A02	N = isolate E03	T = isolate C13		
	1045	F												S = isolate Cq-02				
	1046	A												V = isolate C11				
	1047	S												Y = isolate A06				
	1051	Q											L	R = isolate F01				
	1052	S					L							T = isolates D09,				

Table 26 continued overleaf

RF3	No	aa	A	A	B	B	B*	C	Cq	D	D	E	F	value 2	value 3
	1055	Q					K						K		
	1058	P												Q = isolate F02	
	1059	I			L	L	L			L	L		V	L = isolate E03	F = isolate B'04
	1060	S										T		T = isolate D09	S = isolate E03
	1065	D								N	N			E = isolate A09	D = isolate
	1066	T	S	S				S	S						

* Quasi-species change within a genotype or subgroup is listed in the right hand columns labelled value 2 etc. together with the codes of the isolates displaying this variation.

APPENDIX E

MEAN PEAK COUNTS OF ^{3}H-THYMIDINE INCORPORATED INTO PERIPHERAL BLOOD MONONUCLEAR CELLS IN RESPONSE TO MITOGEN AND HBV CORE ANTIGEN IN ACUTE AND CHRONIC HEPATITIS

Table 27. (a) Preliminary counts used to determine optimum concentration of PHA and the effect of peptide solvents on cell proliferation. (b) Mean counts of ^{3}H-thymidine incorporated into PBMCs in response to stimulation by mitogen or overlapping core antigen peptides in acute (patients ES06, ES15, ES16 and HB05) and chronic hepatitis (patients N231, N232, N234, N235, N236, N237 and N238)

(a)

SOLVENT	ANTIGEN	CONC	MEAN	SD	SE	95 % CI	SI
NONE	NONE	-	101.00	22.99	11.49	22.53	1.00
Ammonia	NONE	-	109.88	31.82	15.91	31.19	1.09
Acetic acid	NONE	-	112.50	25.12	12.56	24.62	1.11
DMSO	NONE		141.38	109.46	54.73	107.27	1.40
NONE	PHA	2.50	18629.75	6659.81	3329.91	6526.61	184.45
NONE	PHA	5.00	41672.25	2339.13	1169.57	2292.35	412.60
NONE	PHA	10.00	39557.50	4579.50	2289.75	4487.91	391.66
Ammonia	PHA	10.00	40189.88	10485.09	5242.55	10275.39	397.92
Acetic acid	PHA	10.00	44762.63	8137.43	4068.72	7974.68	443.19
DMSO	PHA	10.00	35902.00	6043.70	3021.85	5922.83	355.47
NONE	PHA	20.00	43873.00	1935.27	967.64	1896.56	434.39

(b)

SPECIMEN	ANTIGEN	CONC	MEAN PEAK RESPONSE	SD	SE	95 % CI	SI
HB05	PHA	10.00	8756.13	1785.64	892.82	1749.93	61.34
	NONE	0.00	142.75	25.87	14.93	29.27	1.00
	p22-4	2.00	348.00	65.64	37.90	74.28	2.44
	p22-5	2.00	962.00	280.01	198.00	388.08	6.74
	p22.6	0.01	381.00	85.34	39.25	76.93	2.67
	p22-7	0.01	560.50	7.07	5.00	9.80	3.93
	p22-8	0.01	241.33	20.86	14.75	28.91	1.69
	p22-9	0.01	304.25	60.76	30.38	59.55	2.13
	p22-10	2.00	682.83	152.14	87.84	172.16	4.78
	p22-11	2.00	190.50	63.79	36.83	72.19	1.33
	p22-12	0.01	289.38	103.48	51.74	101.41	2.03
	p22-13	2.00	299.33	54.56	31.50	61.74	2.10
	p22-14	2.00	388.17	101.56	58.64	114.93	2.72
	p22-15	0.01	294.33	145.23	83.85	164.34	2.06
	p22-16	2.00	129.00	22.25	12.85	25.18	0.90
	p22-17	0.01	1014.00	256.68	181.50	355.74	7.10
	p22-18	2.00	341.25	42.07	29.75	58.31	2.39
	p22-19	2.00	234.00	74.95	53.00	103.88	1.64
	p22-20	2.00	295.50	12.02	8.50	16.66	2.07

SPECIMEN	ANTIGEN	CONC	MEAN PEAK RESPONSE	SD	SE	95 % CI	SI
ES06	PHA	10.00	41715.50	2837.62	2006.50	3932.74	350.55
	NONE	-	119.00	4.95	3.50	6.86	1.00
	p22-4	0.25	855.11	385.94	272.90	534.88	7.19
	p22-5	0.25	332.25	158.75	112.25	220.01	2.79
	p22-6	0.25	255.25	23.69	16.75	32.83	2.14
	p22-7	0.25	322.75	176.42	124.75	244.51	2.71
	p22-8	0.25	284.50	129.40	91.50	179.34	2.39
	p22-9	0.25	342.75	66.82	47.25	92.61	2.88
	p22-10	0.25	246.50	70.00	49.50	97.02	2.07
	p22-11	0.25	179.50	19.80	14.00	27.44	1.51
	p22-12	0.25	155.75	53.39	37.75	73.99	1.31
	p22-13	0.25	197.25	51.27	36.26	71.05	1.62
	p22-14	0.25	325.75	59.04	41.75	81.83	2.39
	p22-15	0.25	165.50	45.25	32.00	62.72	1.39
	p22-16	0.25	161.50	10.61	7.50	14.70	1.36
	p22-17	1.00	141.50	19.80	14.00	27.44	1.62
	p22-18	2.00	161.25	28.64	20.25	39.69	1.36
	p22-19	1.00	106.75	14.50	10.25	20.09	0.90
	p22-20	1.00	153.00	40.31	28.50	55.86	1.29

Table 27(b) continued overleaf

SPECIMEN	ANTIGEN	CONC	MEAN PEAK RESPONSE	SD	SE	95 % CI	SI
ES15	PHA	10	4405.00	227.81	113.91	223.26	21.16
	NONE	0	208.17	16.93	9.77	19.16	1.00
	p22-4	2	573.00	61.34	30.67	60.11	2.57
	p22-5	0.25	529.75	106.14	53.07	104.02	2.37
	p22-6	2	485.83	77.38	44.68	87.57	2.17
	p22-7	2	415.83	64.58	37.29	73.08	1.86
	p22-8	2	435.17	78.65	45.41	89.00	1.95
	p22-9	2	659.75	30.76	21.75	42.63	2.95
	p22-10	0.25	449.63	51.19	25.59	50.16	2.01
	p22-11	0.25	382.17	63.38	36.60	71.73	1.71
	p22-12	0.25	544.50	112.07	64.70	126.82	2.44
	p22-13	0.25	663.75	82.38	58.25	114.17	2.97
	p22-14	2	490.63	94.39	47.20	92.50	2.20
	p22-15	0.25	485.63	103.56	51.78	101.49	2.17
	p22-16	0.25	581.75	86.77	43.38	85.03	2.60
	p22-17	0.01	738.00	57.28	40.50	79.38	3.30
	p22-18	2	559.33	191.03	110.29	216.17	2.50
	p22-19	2	458.50	76.50	38.25	74.97	2.05
	p22-20	2	673.25	103.59	73.25	143.57	3.01

SPECIMEN	ANTIGEN	CONC	MEAN PEAK RESPONSE	SD	SE	95 % CI	SI
ES16	PHA	10.00	3821.99	936.82	468.41	918.08	29.00
	NONE	0.00	131.78	3.29	2.33	4.56	1.00
	p22-4	2.00	245.20	24.62	14.21	27.85	1.86
	p22-5	2.00	235.00	35.01	20.21	39.62	1.78
	p22.6	2.00	297.81	113.16	56.58	110.90	2.26
	p22-7	0.01	262.25	36.59	21.13	41.41	1.99
	p22-8	0.01	191.18	27.90	19.73	38.66	1.45
	p22-9	2.00	371.53	20.33	14.38	28.18	2.82
	p22-10	2.00	281.35	133.33	94.21	184.66	2.14
	p22-11	0.01	182.56	38.49	19.25	37.72	1.39
	p22-12	2.00	171.75	74.31	37.15	72.82	1.30
	p22-13	2.00	272.20	103.55	51.78	101.48	2.07
	p22-14	2.00	309.37	59.03	34.08	66.80	2.35
	p22-15	2.00	298.45	174.40	32.95	64.58	2.26
	p22-16	0.01	190.70	62.47	36.07	70.69	1.45
	p22-17	2.00	341.98	31.42	18.14	35.56	2.60
	p22-18	2.00	219.21	44.24	22.12	43.36	1.66
	p22-19	2.00	147.69	42.53	21.26	41.68	1.12
	p22-20	0.01	141.56	48.85	28.05	54.97	1.07

Table 27(b) continued overleaf

SPECIMEN	ANTIGEN	CONC	MEAN PEAK RESPONSE	SD	SE	95 % CI	SI
N231	PHA	10.00	3257.63	246.99	123.50	242.05	11.27
	NONE	0.00	289.00	23.40	11.70	22.93	1.00
	p22-4	2.00	805.33	301.47	174.05	341.15	2.79
	p22-5	2.00	632.33	146.49	84.57	165.77	2.19
	p22.6	2.00	509.33	237.88	137.34	269.19	1.76
	p22-7	0.01	568.67	113.10	65.30	127.98	1.97
	p22-8	0.01	284.50	153.65	41.07	80.49	0.98
	p22-9	0.01	474.00	273.78	8.00	15.68	1.64
	p22-10	2.00	303.50	77.09	38.55	75.55	1.05
	p22-11	0.01	479.50	151.88	75.94	148.84	1.66
	p22-12	0.01	305.88	75.34	37.67	73.84	1.06
	p22-13	2.00	339.88	60.45	30.23	59.24	1.18
	p22-14	2.00	331.50	65.15	32.58	63.85	1.15
	p22-15	2.00	354.25	62.28	31.14	61.03	1.23
	p22-16	0.01	286.33	10.30	5.95	11.66	0.99
	p22-17	0.01	291.50	126.97	73.30	143.68	1.01
	p22-18	0.01	253.67	42.14	24.33	47.69	0.88
	p22-19	2.00	300.83	71.00	40.99	80.34	1.04
	p22-20	0.01	307.75	87.12	43.56	85.38	1.06

SPECIMEN	ANTIGEN	CONC	MEAN PEAK RESPONSE	SD	SE	95 % CI	SI
N232	PHA	10.00	3352.00	199.44	115.14	225.68	12.60
	NONE	0.00	266.00	97.25	56.15	110.05	1.00
	p22-4	0.01	411.00	41.72	29.50	57.82	1.55
	p22-5	2.00	464.33	86.63	50.02	98.03	1.75
	p22.6	0.01	413.25	42.07	29.75	58.31	1.55
	p22-7	2.00	351.50	109.55	63.25	123.97	1.32
	p22-8	2.00	514.67	143.26	82.71	162.12	1.93
	p22-9	2.00	357.50	22.52	13.00	25.48	1.34
	p22-10	2.00	597.33	180.06	103.96	203.76	2.25
	p22-11	0.01	455.00	41.01	29.00	56.84	1.71
	p22-12	2.00	376.00	68.14	39.34	77.11	1.41
	p22-13	0.01	289.50	49.40	28.52	55.91	1.09
	p22-14	2.00	618.50	211.42	149.50	293.02	2.33
	p22-15	0.01	602.00	26.92	15.54	30.46	2.26
	p22-16	2.00	343.28	64.39	37.18	72.87	1.29
	p22-17	2.00	789.25	178.54	126.25	247.45	2.97
	p22-18	2.00	715.33	82.77	47.79	93.66	2.69
	p22-19	2.00	655.75	51.97	36.75	72.03	2.47
	p22-20	0.01	359.75	107.83	76.25	149.45	1.35

Table 27(b) continued overleaf

SPECIMEN	ANTIGEN	CONC	MEAN PEAK RESPONSE	SD	SE	95 % CI	SI
N234	PHA	10.00	5025.88	374.89	187.44	367.39	18.94
	NONE	0.00	265.33	57.42	33.15	64.98	1.00
	p22-4	2.00	575.33	142.88	82.49	161.69	2.17
	p22-5	0.01	446.50	44.74	22.37	43.85	1.68
	p22.6	2.00	427.33	68.16	39.35	77.13	1.61
	p22-7	2.00	318.83	57.23	33.04	64.76	1.20
	p22-8	0.01	548.00	72.12	51.00	99.96	2.07
	p22-9	0.01	328.67	46.76	27.00	52.91	1.24
	p22-10	0.01	364.00	130.68	75.45	147.87	1.37
	p22-11	0.01	313.75	8.13	5.75	11.27	1.18
	p22-12	0.01	343.00	83.44	59.00	115.64	1.29
	p22-13	0.01	355.00	29.35	14.68	28.76	1.34
	p22-14	0.01	674.00	211.00	114.71	224.83	2.54
	p22-15	0.01	426.50	142.64	82.35	161.42	1.61
	p22-16	0.01	356.67	40.82	20.41	40.01	1.34
	p22-17	2.00	513.50	349.33	201.69	395.31	1.94
	p22-18	2.00	360.50	45.91	26.51	51.95	1.36
	p22-19	0.01	447.17	214.86	124.05	243.14	1.69
	p22-20	2.00	461.00	4.95	3.50	6.86	1.74

SPECIMEN	ANTIGEN	CONC	MEAN PEAK RESPONSE	SD	SE	95 % CI	SI
N235	PHA	10.00	8390.67	416.35	240.38	471.15	38.37
	NONE	0.00	218.67	65.11	37.59	73.68	1.00
	p22-4	2.00	512.00	132.23	93.50	183.26	2.34
	p22-5	0.01	207.33	85.45	49.33	96.69	0.95
	p22.6	2.00	195.25	20.15	14.25	27.93	0.89
	p22-7	2.00	519.00	69.50	40.13	78.65	2.37
	p22-8	0.01	251.00	13.50	7.79	15.28	1.15
	p22-9	2.00	196.00	51.62	36.50	71.54	0.90
	p22-10						
	p22-11	2.00	305.00	61.92	35.75	70.07	0.60
	p22-12	0.01	185.00	34.65	24.50	48.02	0.85
	p22-13	0.01	268.17	72.87	42.07	82.46	1.23
	p22-14	0.01	299.00	209.30	148.00	290.08	1.37
	p22-5						
	p22-16	2.00	335.25	29.34	20.75	40.67	1.53
	p22-17	2.00	233.00	137.89	97.50	191.10	1.07
	p22-18	0.01	249.00	69.30	49.00	96.04	1.14
	p22-19	0.01	177.17	75.79	43.76	85.76	0.81
	p22-20	0.01	310.50	99.70	70.50	138.18	1.42

Table 27(b) continued overleaf

SPECIMEN	ANTIGEN	CONC	MEAN PEAK RESPONSE	SD	SE	95 % CI	SI
N236	PHA	10.00	3831.25	199.43	99.71	195.44	5.61
	NONE	0.00	683.50	97.92	56.53	110.81	1.00
	p22-4	2.00	623.33	266.61	153.93	301.70	0.91
	p22-5	2.00	821.50	148.49	105.00	205.80	1.20
	p22.6	0.01	669.88	160.29	80.15	157.08	0.98
	p22-7	0.01	683.17	117.85	68.04	133.36	1.00
	p22-8	0.01	730.88	146.78	73.39	143.84	1.07
	p22-9	0.01	758.00	161.57	93.28	182.83	1.11
	p22-10	0.01	790.50	70.71	50.00	98.00	1.16
	p22-11	2.00	770.00	170.21	98.27	192.62	1.13
	p22-12	2.00	656.00	85.46	49.34	96.70	0.96
	p22-13	0.01	713.75	131.17	92.75	181.79	1.04
	p22-14	0.01	985.33	235.40	135.91	266.38	1.44
	p22-15	0.01	816.63	117.92	58.96	115.56	1.19
	p22-16	2.00	704.13	252.37	126.18	247.32	1.03
	p22-17	0.01	714.83	133.53	77.10	151.11	1.05
	p22-18	0.01	686.33	31.07	17.94	35.16	1.00
	p22-19	0.01	567.00	179.71	103.76	203.37	1.61
	p22-20	2.00	749.83	131.61	75.99	148.94	1.10

SPECIMEN	ANTIGEN	CONC	MEAN PEAK RESPONSE	SD	SE	95 % CI	SI
N237	PHA	10.00	6360.33	742.10	428.45	839.77	5.27
	NONE	0.00	1206.50	117.74	58.87	115.39	1.00
	p22-4	2.00	1254.67	313.54	181.03	354.81	1.04
	p22-5	2.00	1055.63	184.20	130.25	255.29	0.87
	p22.6	2.00	1130.67	180.41	104.16	204.15	0.18
	p22-7	0.01	1055.75	23.69	16.75	32.83	0.88
	p22-8	0.01	1032.83	170.96	98.70	193.46	0.86
	p22-9	0.01	1207.00	50.08	28.92	56.67	1.00
	p22-10	2.00	1108.50	215.01	124.13	243.30	0.92
	p22-11	2.00	1306.50	330.93	149.69	293.40	1.08
	p22-12	0.01	1042.50				0.86
	p22-13	0.01	990.63	511.30	255.65	501.07	0.82
	p22-14	0.01	1555.88	351.68	175.84	344.64	1.29
	p22-15	2.00	861.00	57.28	40.50	79.38	0.71
	p22-16	0.01	980.33	76.54	44.19	86.61	0.81
	p22-17	2.00	1394.38	257.85	128.92	252.69	1.16
	p22-18	0.01	1434.00	603.53	301.77	591.46	1.19
	p22-19	2.00	1183.75	73.19	51.75	101.43	0.98
	p22-20	2.00	2052.25	352.24	176.12	345.19	1.70

Table 27(b) continued overleaf

SPECIMEN	ANTIGEN	CONC	MEAN PEAK RESPONSE	SD	SE	95 % CI	SI
N238	PHA	10.00	6334.67	146.58	84.63	165.87	38.82
	NONE	0.00	163.17	31.91	18.42	36.11	1.00
	p22-4	2.00	228.17	97.33	56.19	110.14	1.40
	p22-5	0.01	189.00	27.53	15.89	31.15	1.16
	p22.6	2.00	170.50	59.44	34.32	67.26	1.04
	p22-7	2.00	268.25	105.01	74.25	145.53	1.64
	p22-8	2.00	247.50	46.67	33.00	64.68	1.52
	p22-9	0.01	348.25	32.17	22.75	44.59	2.13
	p22-10						
	p22-11	0.01	228.25	41.37	29.25	57.33	1.40
	p22-12	0.01	156.75	12.37	8.75	17.15	0.96
	p22-13	0.01	135.75	29.34	20.75	40.67	0.83
	p22-14	0.01	127.00	29.70	21.00	41.16	0.78
	p22-15	0.01	170.00	4.95	3.50	6.86	1.04
	p22-16	2.00	186.00	106.77	75.50	147.98	1.14
	p22-17	0.01	326.75	206.83	146.25	286.65	2.00
	p22-18	0.01	157.00	0.71	0.50	0.98	0.96
	p22-19	2.00	140.17	17.27	9.97	19.54	0.86
	p22-20	2.00	194.83	80.93	46.73	91.59	1.19

DMSO = dimethyl sulphoxide
SD = standard deviation SE = standard error 95 % CI = 95 percent confidence interval
SI = stimulation index

APPENDIX F

SUMMARY OF SOURCE, SIZE AND SIGNIFICANCE OF SAMPLES USED IN EACH OF THE SEVEN SECTIONS OF THE STUDY

F.1 Molecular epidemiology of HBV infection

Thirty-seven urban South African specimens were selected from HBsAg-positive specimens sent to the NIV for routine diagnostic serology. The rural South African specimens consisted of children visiting the primary health clinics in the rural township of Mdantsane in the Eastern Cape Province. Children were included in the rural study if there was no history of hepatitis B immunization. Fifty-seven HBsAg-positive samples were randomly selected from this cohort and 37 of these had sufficient volume remaining, as well as sufficient HBV DNA to enable PCR amplification, and were used for molecular characterization of the circulating genotypes of HBV (Vardas, 1999).

The 49 genotyped specimens from the northern provinces of rural Namibia included 16 specimens from the pilot HBV prevalence study of the northern provinces, 25 selected specimens from a 1984 Ovamboland study of Botha *et al.* and eight specimens from chronic carriers selected for the proliferation study which is described in Chapter 9.

F.1.1 Calculation of suitable sample size

A suitable sample size can be calculated using the formula:

$Z^2 p^{(1-p)}/_d{}^2$

where

d = the acceptable margin of error

Z = the acceptable confidence level

p = the expected proportion of responses

Since 1.96 standard deviations contain 95 percent of the values on a normal curve, using the traditionally recognized acceptable risk of 95 percent, Z is equal to 1.96. Taking the average proportion of HBsAg-positive individuals in southern Africa as 10 percent, p is equal to 0.1.

Accepting a confidence interval of 5 percent on either side of the result makes d equal to 0.1.

Thus acceptable sample size is $(1.96)^2 0.1^{(1-0.1)}/_{0.1^2}$

$3.84 \times 0.1 \times {}^{(0.9)}/_{0.01}$

$\underline{34.56}$

Thus urban and rural samples of 37 each were used .

A total of 48 specimens from the rural northern provinces of Namibia were also tested.

F.1.2 Ninety-five percent confidence interval for each population proportion

The 95 percent confidence interval for a population proportion is:

$[p-1.96\sqrt{p(1-p)}/_n,\ p+1.96\sqrt{p(1-p)}/_n]$

where

p is the observed proportion

and n is the sample size (Petrie, 1987).

Using this formula each proportion calculated in Chapter 4 is given in Table 28(a) together with its confidence interval (shown in brackets) based on relevant sample size.

F.2 Sequence characterization of subgroup A'

HBV DNA was sequenced over the preS2 and S genes for 29 South African patients, 11 of whom had acute HBV disease and 18 of whom were chronic carriers. This study was performed to establish the sequence differences in local specimens which were giving anomalous results in the RFLP analysis. The frequency of the genotypes and subtypes found in this subset of specimens (Chapter 5) agreed (within calculated confidence intervals) with those found in the larger study outlined in Chapter 4 and summarized in Table 28a (Table 28b).

F.3 Hotspots on the HBV genome and the relationship between the genotypic groups of HBV

The third study examined the 'hotspots' on the HBV genome and is reported in Chapter 6 while the fourth study examined the relationship between the genotypic groups of HBV and is reported in Chapter 7. Both these studies used the same 65 GenBank specimens listed in paragraph 6.2.1. The only selection criteria was that the full-genome had been sequenced and these specimens represent all the full-genomes available from GenBank at the time. The

Table 28. Summary of source, size and significance of samples used in (a) Determination of molecular epidemiology of the different genotypes of HBV in rural and urban South Africa and the rural areas of Northern Namibia and (b) Molecular epidemiology of a subset of 29 South African specimens genotyped by sequence analysis

(a)

Country	urban/rural	Source	N	HBsAg +ve	Selection Strategy	No. tested	Result (95 % CI for population proportion)
South Africa	urban Gauteng	40 to 50 000 diagnostic specimens are sent to the NIV annually from Doctors, hospitals and Clinics in the Gauteng Province of which 8 to 9 000 specifically request hepatitis studies	± 50 000	± 5,000	random sampling of NIV stored HBsAg-positive diagnostic specimens	37	genotype A 69 6 (54 1 84 0) % [subgroup A 25 0 (11 1 39 0) % subgroup A' 44 6 (28 6 60 6) %] genotype D 16 1 (4 1 28.1) % genotype C 7 1 (0.0 15 4) % genotype B 3 6 (0 0 9 6) % genotype E 3 6 (0 0 9 6) %
	rural Eastern Province	Age stratified sample of children 0-6 years of age from the Eastern Cape Province of South Africa	2299	238	random sampling	37	genotype A 89 2 (79 2 99 2) % [subgroup A 2 7 (0 7 9) % subgroup A' 86 5 (75 5 97 5) %] genotype D 10 8 (0.7 20 8) %
Namibia	rural NW and NE province	High school children, all first time blood donors from the NW and NE provinces of Namibia	230	39	all specimens with sufficient volume tested	16	genotype E 93 8 (87 0 100 6) % genotype D 6 2 (0 0 13 0) %
	rural Ovamboland in NW province	Existing cohort from a previous study collected from State Hospital and regional clinic patients in Ovamboland in NE Namibia	1338 children 1272 mothers	201 children 140 mothers	random sampling of HBsAg- and HBeAg-positive mothers and children	24	
	rural Rundu NE province	Chronic carriers identified among first time blood donors	8	8	all specimens tested	8	

(b)

Country	urban/ rural	Source	N	HBsAg +ve	Selection Strategy	No. tested	Result (95 % CI for population proportion)
South Africa	urban Gauteng	HBV DNA extracted and amplified from 11 acute hepatitis B and 18 chronic carriers from the Gauteng province of South Africa were sequenced and compared with 81 global sequences from GenBank	29	29	all specimens tested	29	genotype A 82 7 (68 9 96 5) % [subgroup A 23 6 (8 1 39 0) % subgroup A' 59.1 (41 2 77 0) %] genotype D 10.3 (0 0 21 4) % genotype B 3 4 (0.0 6 6) % genotype C 3.4 (0 0 6 6) %

N is the number of specimens collected in each cohort described under 'Source'
HBsAg-+ve is the number of surface antigen specimens in the original cohort and
No. tested is the number of specimens with sufficient volume and HBV DNA (to be amplified by PCR) and therefore suitable for genotype determination
CI is the 95 percent confidence interval of each proportion

results of these studies may therefore not show a truly average cross section of the similarities and differences between the strains of HBV in circulation in any particular population since scientists are more likely to sequence virus isolates from clinically interesting cases. Table 26, Appendix D, lists all variation found within the database of 65 GenBank specimens. Mosaic structure was demonstrated statistically in 11/65 (16.9 percent) of the 65 isolates using bootstrap resampling.

F.4 HLA frequency and the susceptibility to or protection from HBV in southern Africa and mapping of the immunogenic regions of the core gene

We performed class II HLA-DR typing on 11 patients and two healthy individuals (listed in Table 18) and compared our results with published frequency analyses (listed in Table 16). We also reviewed these results with respect to published studies which reported the effect of different HLA profiles on the susceptibility to or protection from HBV infection (Table 19). These results are reported in Chapter 8. The typing results from the 13 specimens were not intended to be representative of frequencies in the southern African population but were performed so that we could examine peptides found to be immunogenic in the patients for possible, relevant epitopes. These results are reported in Chapter 9.

F.5 HBV core gene variation

Finally core gene sequence data from 14 southern African patients were compared with 161 core gene sequences from GenBank. Differences in the sequences of these 175 specimens are shown in the various graphs of Chapter 10.

REFERENCES:

Abdool Karim, S.S.A., Coovadia, H.M., Windsor, I.M., Thejpal, R., Van den Ende, J. and Fouche, A. (1988). The prevalence and transmission of hepatitis B virus infection in urban, rural and institutionalized black children of Natal/KwaZulu, South Africa. *Int. J. Epid.;* **17:** 168-173.

Akarca, U.S. and Lok, A.S.F. (1995). Naturally occurring hepatitis B virus core gene mutations. *Hepatology;* **22:** 50-60.

Allsopp, C.E., Harding, R.M., Taylor, C., Bunce, M, Kwiatkowski, D., Anstey, N., Brewster, D., McMichael, A.J., Greenwood, B.M. and Hill, A.V. (1992). Interethnic genetic differentiation in Africa: HLA class I antigens in The Gambia. *Am. J. Hum. Genet.*; **50:** 411-421.

Almarri, A. and Batchelor, J.R., 1994. HLA and hepatitis B infection. *The Lancet*; **344:** 1194-1195.

Antoni, B.A., Rodriguez-Crespo, I., Gomez-Gutierrez, J., Peterson, D. and Gavilanes, F. (1994) Site-directed mutagenesis of cysteine residues of hepatitis B surface antigen: Analysis of two single mutants and the double mutant. *Euro. J. Biochem.;* **222:** 121-127.

Arauz-Ruiz, P., Norder, H., Visona, K.A. and Magnius, L.O. (1997a). Genotype F prevails in HBV infected patients of hispanic origin in central America and may carry the preCore stop mutant. *J. Med. Virol.*; **51:** 305-312.

Arauz-Ruiz, P., Norder, H., Visona, K.A. and Magnius, L.O. (1997b). Molecular epidemiology of hepatitis B virus in Central America reflected in the genetic variablility of the small S gene. *J. Infect. Dis.;* **176:** 851-858.

Arii, M., Takada, S. and Koike, K. (1992). Identification of three essential regions of hepatitis B virus X protein for *trans* activation function. *Oncogene;* **7:** 397-403.

Ashton-Rickardt, P.G. and Murray, K. (1989). Mutants of the hepatitis B virus surface antigen that define some antigenically essential residues in the immunodominant *a* region. *J. Med. Virol.;* **29:** 196-203.

Aspinall, S., Joubert, J.J., Evans, A.C., Joseph, S., Steele, A.D. and Lecatsas, G. (1994). Prevalence of hepatitis B in !Kung (San) children from Bushmanland, Namibia. *Annals Trop. Paed.*; **14:** 163-167.

Bahn, A., Gerner, P., Martin,, U., Bortolotti, F. and Wirth, S. (1997). Detection of different viral strains of hepatitis B virus in chronically infected children after seroconversion from HBsAg to anti-HBs indicating viral persistence. *J. Hepatology*; **27:** 973-978.

Baptista, M., Kramvis, A. and Kew, M.C. (1999). High prevalence of 1762(T) 1764(A) mutations in the basic core promoter of hepatitis B virus isolated from black Africans with hepatocellular carcinoma compared with asymptomatic carriers. ***Hepatology;* 29:** 946-953.

Barnaba, V., Franco, A., Alberti, A., Balsano, C., Benvenuto, R. and Balsano, F.(1989). Recognition of hepatitis B virus envelope proteins by liver-infiltrating T lymphocytes in chronic HBV infection. *J. Immunol.;* **143:** 2650-2655.

Barnaba, V., Franco, A., Paroli, M., Benvenuto, R., De Petrillo, G., Burgio, V.L., Santilio, I., Balsano, C., Bonavita, M.S., Cappelli, G., Colizzi, V., Cutrona, G. and Ferrarini, M. (1994). Selective expansion of cytotoxic T lymphocytes with a CD4+CD56+ surface phenotype and a T helper type I profile of cytokine secretion in the liver of patients chronically infected with hepatitis B virus. *J. Immunol.;* **152:** 3074-3087.

Baumert, T.F., Rogers, S.A., Hasegawa, K. and Liang, T.J. (1996). Two core promoter mutations identified in a hepatitis B virus strain associated with fulminant hepatitis result in enhanced viral replication. *J. Clin. Invest.;* **98:** 2268-2276.

Beasley, R.P., Hwang, L.Y., Lin, C.C. and Chien, C.S. (1981). Hepatocellular carcinoma

and hepatitis B virus. A prospective study of 22707 men in Taiwan. *Lancet;* **2:** 1129-1133.

Beasley, R.P. and Hwang, L.-Y. (1984). Epidemiology of hepatocellular carcinoma. In: *Viral hepatitis and liver disease* pp.209-224. (Eds.: Vyas, G.N., Dienstag, J.L., Hoofnagle, J.H.). Orlando: Grune and Stratton.

Benn, J. and Schneider, R.J. (1994). Hepatitis B virus HBx protein activates *Ras*-GTP complex formation and establishes a *Ras*, *Raf*, MAP kinase signaling cascade. *Proc. Natl. Acad. Sci. USA*; **91:** 10350-10354.

Benn, J. and Schneider, R.J. (1995). Hepatitis B virus HBx protein deregulates cell cycle checkpoint controls. *Proc. Natl. Acad. Sci. USA;* **92:** 11215-11219.

Benn, J., Su, F., Doria, M. and Schneider, R.J. (1996). Hepatitis B virus HBx protein induces transcription factor AP-1 by activation of extracellular signal-regulated and c-*Jun* N-terminal mitogen-activated protein kinases. *J. Virol.;* **70:** 4978-4985.

Bertherat, E., Georges-Courbot, M.C., Nabias, R., Georges, A.J. and Renaut, A. (1998). Seroprevalence of four sexually transmitted diseases in a semi-urban population of Gabon. *Int. J. STD AIDS;* **9:** 31-36.

Berting, A., Hahnen, J., Kröger, M. and Gerlich,W.H. (1995). Computer-aided studies on the spatial structure of the small hepatitis B surface protein. *Intervirology*; **38:** 8-15.

Bertoletti, A., Chisari, F.V., Penna, A., Guilhot, S., Galati, L., Missale, G., Fowler, P., Schlicht, H.-J., Vitiello, A., Chesnut, R.C., Fiaccadori, F. and Ferrari, C. (1993). Definition of a minimal optimal cytotoxic T-cell epitope within the hepatitis B virus nucleocapsid protein. *J. Virol.;* **67:** 2376-2380.

Bertoletti, A., Sette, A., Chisari, F.V., Penna, A., Levrero, M., De Carli, M., Flaccadori, F., and Ferrari, C. (1994). Natural variants of cytotoxic epitopes are T-cell receptor antagonists for antiviral cytotoxic T cells. *Nature;* **369:** 407-410.

Bertoletti, A., Southwood, S., Chesnut, R., Sette, A., Falco, M., Ferrara, G.B., Penna, A., Boni, C., Fiaccadori, F. and Ferrari, C. (1997). Molecular features of HBV nucleocapsid T-cell epitope 18-27: Interaction with HLA and TCR. *Hepatology*; **26:** 1027-1034.

Blitz, L., Pujol, F.H., Swenson, P.D., Porto, L., Atencio, R., Araujo, M., Costa, L., Monsalve, D.C., Torres, J.R., Fields, H.A., Lambert, S., Van Geyt, C., Norder, H., Magnius, L.O., Echevarría, J.M. and Stuyver, L. (1998). Antigenic diversity of hepatitis B virus strains of genotype F in Amerindians and other population groups from Venezuela. *J. Clin. Micro.*; **36:** 648-651.

Blum, H.E., Galun, E., Liang, T.J., von Weizsäcker, F. and Wands, J.R. (1991). Naturally occurring missense mutation in the polymerase gene terminating hepatitis B virus replication. *J. Virol.*; **65:** 1836-1842.

Bock, C.-T., Schwinn, S., Schroder, C.H., Velhagen, I. and Zentgraf, H. (1996). Localization of hepatitis B virus core protein and viral DNA at the nuclear membrane. *Virus Genes;* **12:** 53-63.

Boisier, P., Rabarijaona, L., Piollet, M., Roux, J.F. and Zeller, H.G. (1996). Hepatitis B virus infection in general population in Madagascar: Evidence for different epidemiological patterns in urban and rural areas. *Epid. Inf.*; **117:** 133-137.

Bollyky, P.L., Rambaut, A., Harvey, P.H. and Holmes, E.C. (1996). Recombination between sequences of hepatitis B virus from different genotypes. *J. Mol. Evol;*. **42:** 97-102.

Bollyky, P.L. and Holmes, E.C. (1999). Reconstructing the complex evolutionary history of hepatitis B virus. *J. Mol. Evol.*; **49:** 130-141.

Botha, J.F., Dusheiko, G.M., Ritchie, M.J.J., Mouton, H.W.K. and Kew, M.C. (1984). Hepatitis B virus carrier state in black children in Ovamboland: Role of perinatal and horizontal infection. *The Lancet;* **1:** 1210-1212.

Bottcher, B., Tsuji, N., Takahashi, H., Dyson, M.R., Zhao, S., Crowther, R.A. and Murray, K. (1998). Peptides that block hepatitis B virus assembly: analysis by cryomicroscopy, mutagenesis and transfection. *The EMBO Journal*; **17**: 6839-6845.

Bowyer, S.M., van Staden, L., Kew, M.C. and Sim, J.G.M. (1997). A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. *J. Gen. Virol.*; **78:** 1719-1729.

Bowyer, S.M. and Sim, J.G.M. (2000). Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates. *J. Gen. Virol.*; **81:** 379-392.

Bressac, B., Kew, M., Wands, J. and Ozturk, M. (1991). Selective G to T mutations of p53 gene in Hepatocellular Carcinoma from southern Africa. *Nature;* **350:** 429-431.

Bruss, V. and Ganem, D. (1991). Mutational analysis of hepatitis B surface antigen particle assembly and secretion. *J. Virol.;* **65:** 3813-3820.

Bruss, V., Lu, X., Thomssen, R. and Gerlich, W.H. (1994). Post-translational alterations in transmembrane topology of the hepatitis B virus large envelope protein. *EMBO;* **13:** 2273-2279.

Budkowska, A., Bedossa, P., Groh, F., Louise, A. and Pillot, J. (1995). Fibronectin of human liver sinusoids binds hepatitis B virus: identification by an anti-idiotypic antibody bearing the internal image of the pre-S2 domain. *J Virol.*; **69:** 840-848.

Buscher, M., Reiser, W., Will, H. and Schaller, H. (1985). Transcripts and the putative RNA pregenome of duck hepatitis B virus: implications for reverse transcription. *Cell;* **40:** 717-724.

Carman, W.F., Zanetti, A.R., Karayiannis, P., Waters, J., Manzillo, G., Tanzi, E.,

Zuckerman, A.J. and Thomas, H.C. (1990). Vaccine-induced escape mutant of hepatitis B virus. *The Lancet;* **336:** 325-329.

Carman, W.F., Thursz, M., Hadziyannis, S., McIntyre, G., Colman, K., Gioustoz, A., Fattovich, G., Alberti, A. and Thomas, H.C. (1995). Hepatitis B e antigen negative chronic active hepatitis: hepatitis B virus core mutations occur predominantly in known antigenic determinants. *J. Viral Hepat.*; **2:** 77-84.

Carman, W.F., Boner, W., Fattovich, G., Colman, K., Dorman, E.S., Thursz, M. and Hadziyannis, S. (1997). Hepatitis B virus core protein mutations are concentrated in B cell epitopes in progressive disease and in T helper cell epitopes during clinical remission. *J. Infect. Dis.*; **175:** 1093-1100.

Caselmann, W.H. (1996). *Trans*-activation of cellular genes by hepatitis B virus proteins: A possible mechanism of hepatocarcinogenesis. In: *Advances in viral research.* 1st ed. Vol. 47 pp. 253-302. (Eds.: Maramorosch, K., Shatkin, A.J. and Murphy, F.A.). San Diego: Academic Press.

Cattaneo, R., Will, H., Hernandez, N. and Schaller, H. (1983). Signals regulating hepatitis B surface antigen transcription. *Nature*; **305:** 336-338.

Celis, E., Ou, D. and Otvos, L. (1988). Recognition of hepatitis B surface antigen by human T lymphocytes. Proliferative and cytotoxic responses to a major antigenic determinant defined by synthetic peptides. *J. Immunol.;* **140:** 1808-1815.

Cenac, A., Pedroso, M.L., Djibo, A., Develoux, M., Pichoud, C., Lamothe, F., Trepo, C. and Warter, A. (1995). Hepatitis B, C and D virus infections in patients with chronic hepatitis, cirrhosis and hepatocellular carcinoma: a comparative study in Niger. *Am. J. Trop. Med. Hyg.;* **52:** 293-296.

Chen, H.- S.; Kaneko,.S.; Girones, R.; Anderson, R.W.; Hornbuckle, W.E.; Tennant, B.C.; Cote, P.J.; Gerin, J.L.; Purcell, R.H. and Miller, R.H. (1993). The woodchuck hepatitis virus

X gene is important for establishment of virus infection in woodchucks. *J. Virol.*; **67:** 1218-1226.

Chen, M. and Ou, J.-H. (1995). Cell type-dependent regulation of the activity of the negative regulatory element of the hepatitis B virus core promoter. *Virology;* **214:** 198-206.

Chen, P.-J., Chen, C.-R., Sung, J.-L. and Chen, D.-S. (1989). Identification of a doubly spliced viral transcript joining the separated domains for putative protease and reverse transcriptase of hepatitis B virus. *J. Virol.*; **63: 4165-4171**.

Chirara, M.M. and Chetsanga, C.J. (1994). Variant of hepatitis B virus isolated in Zimbabwe. *J. Med. Virol.*; **42:** 73-78.

Chisari, F.V., Klopchin, K., Moriyama, T., Pasquinelli, C., Dunsford, H.A., Sell, S., Pinkert, C.A., Brinster, R.L. and Palmiter, R.D. (1989). Molecular pathogenesis of hepatocellular carcinoma in hepatitis B virus transgenic mice. *Cell;* **59: 1145-1156**.

Chisari, F.V. and Ferrari, A.R. (1995). Hepatitis B virus immunopathogenesis. *Annu. Rev. Immunol.;* **13:** 29-60.

Chung, M.K., Yoon, H., Mun, S.S., Lee, H.G., Kim, Y.J., Lee, T.G., Lim, J.S., Kim, C.M. and Park, S.N. (1999). Induction of cytotoxic T lymphocytes with peptides *in vitro*: identification of candidate T-cell epitopes in hepatitis B virus X antigen. *J. Immunotherapy*; **22:** 279-287.

Colgrove, R., Simon, G. and Ganem, D. (1989). Transcriptional activation of homologous and heterologous genes by the hepatitis B virus X gene product in cells permissive for viral replication. *J. Virol.;* **63:** 4019-4026.

Courouce-Pauty, A.- M. and Soulier, J.P. (1976). HBs antigen subdeterminants as serological entities. *Virology*; **176:** 604-619.

Courouce-Pauty, A.- M., Lemaire, J.M. and Roux, J.F. (1978). New hepatitis B surface antigen subtypes inside the *ad* category. ***Vox Sang.;* 35:** 304-308.

Courouce-Pauty, A.- M., Plançon, A. and Soulier, J.P. (1983). Distribution of HBsAg subtypes in the world. *Vox Sang.*; **44:** 197-211.

Craven, D.E., Zuhier, L.A., Kunches, L.M., Yunis, E.J., Dienstag, J.L., Werner, B.G., Polk, B.F., Snydman, D.R., Platt, R., Crumpacker, C.S., Grady, G.F. and Alper, C.A. (1986). Nonresponsiveness to hepatitis B vaccine in health care workers. *Annals of Int. Med.*; **105:** 356-360.

Dard, P., Sanchez-Mazas, A., Tiercy, J.-M., Magzoub, M., Verduyn, W., Goumaz, C., Jeannet, M., Pellegrini, B., Excoffier, L. and Langaney, A. (1992). HLA-A, -B and -DR differentiation among North and West African populations. In: ***HLA 1991.*** 1st ed. Vol. 1 pp. 632-636. (Eds: Tsuji, K., Aizawa, M. and Sasazuki, T) Oxford: Oxford University Press.

de Meyer, S., Gong, Z.J., Suwandhi, W., van Pelt, J., Soumillion, A. and Yap, S.H. (1997). Organ and species specificity of hepatitis B virus (HBV) infection: a review of literature with a special reference to preferential attachment of HBV to human hepatocytes. *J. Viral Hepat.;* **4:** 145-153.

de Meyer, S., Gong, Z.J., Hertogs, K., Depla, E., van Pelt, J.F., Roskams, T., Maertens, G. and Yap, S.H. (2000). Influence of the administration of human annexin V on *in vitro* binding of small hepatitis B surface antigen to human and rat hepatocytes and on *in vitro* hepatitis B virus infection. *J. Viral Hepat.*; **7:** 104-114.

Delius, H., Gough, N.M., Cameron, C.H. and Murray, K. (1983).Structure of the hepatitis B virus genome. *J. Virol.*; **47:** 337-343.

Desombere, I., Willems, A. and Leroux-Roels, G. (1998). Response to hepatitis B vaccine: multiple HLA genes are involved. *Tissue Antigens*; **51:** 593-604.

Dibisceglie, A.M., Kew, M.C., Dusheiko, G.M., Berger, E.L., Song, E., Paterson, A.C. and Hodkinson, H.J. (1986). Prevalence of hepatitis B virus infection among black children in Soweto. *Br. Med. J.*; **292:** 1440-1442.

Diepolder, H.M., Jung, M.-C., Wierenga, E., Hoffmann, R.M., Zachoval, R., Gerlack, T.J., Scholz, S., Heavner, G., Riethmüller, G. and Pape, G.R. (1996). Anergic TH1 clones specific for hepatitis B virus (HBV) core peptides are inhibitory to other HBV core-specific CD4+ T cells *in vitro*. *J. Virol.*; **70:** 7540-7548.

Diepolder, H.M., Jung, M.-C., Keller, E., Schraut, W., Gerlach, J.T., Gruner, N., Zachooval, R., Hoffmann, R.M., Schirren, C.A., Scholz, S. and Pape, G.R. (1998). A vigorous virus-specific CD4+ T cell response may contribute to the association of HLA-DR13 with viral clearance in hepatitis B. *Clin. Exp. Immunol.*; **113:** 244-251.

Ding, X., Mizokami, M., Yao, G., Xu, B., Orito, E., Ueda, R. and Nakanishi, M. (2001). Hepatitis B virus genotype distribution among chronic hepatitis B virus carriers in Shanghai, China. Intervirol.; **44:** 43-47.

Domingo, E. (1998). Quasispecies and the implications for virus persistence and escape. *Clin. Diag. Virol.*; **10:** 97-101.

Dreyer, K. and Schulte-Holthausen, H. (1991). Casein is a potent enhancer for restriction enzyme activity. *Nucleic Acids Res.*; **19:** 4295.

du Toit, E.D., MacGregor, K.J., Taljaard, D.G. and Oudshoorn, M. (1988). HLA-A, B, C, DR and DQ polymorphisms in three South African population groups: South African Negroes, Cape Coloureds and South African Caucasoids. *Tissue Antigens*; **31:** 109-125.

du Toit, E.D., Taljaard, D.G., Marshall, J., Ritchie, C. and Oudshoorn, M. (1990). The HLA genetic constitution of the Bushmen (San). *Human Immunol.*; **28:** 406-415.

Dusheiko, G.M., Bowyer, S.M., Paterson, A., Song, E., Dibiseglie, A. and Kew, M.C.

(1985). Clinical and serological events accompanying changes in hepatitis B viral replication: case reports. *Liver;* **5:** 77-83.

Dusheiko, G.M., Conradie, J.D., Brink, B.A., Marimuthu, T. and Sher, R. (1989). Differences in the regional prevalence of chronic hepatitis B in southern Africa - implications for vaccination. *S. Afr. Med. J.;* **75:** 473-479.

Eble, B.E., Lingappa, V.R. and Ganem, D. (1990). The N-terminal (pre-S2) domain of a hepatitis B virus surface glycoprotein is translocated across membranes by downstream signal sequences. *J. Virol.;* **64:** 1414-1419.

Edmunds, W.J., Medley, G.F., Nokes, D.J., O'Callaghan, C.J., Whittle, H.C. and Hall, A.J. (1996). Epidemiological patterns of hepatitis B virus (HBV) in highly endemic areas. *Epid. Infect.;* **117:** 313-325.

Enders, G.H., Ganem, D. and Varmus, H. (1985). Mapping the major transcripts of ground squirrel hepatitis virus: the presumptive template for reverse transcriptase is terminally redundant. *Cell;* **41:** 297-308.

Erhardt, A., Reineke, U., Blondin, D., Gerlich, W.H., Adams, O., Heintges, T., Niederau, C. and Haussinger, D. (2000). Mutations of the core promoter and response to interferon treatment in chronic replicative hepatitis B. *Hepatology;* **31:** 716-725.

Estacio, R.C., Chavez, C.C., Okamoto, H., Lingao, A.L., Reyes, M.T., Domingo, E. and Mayumi, M. (1988). Nucleotide sequence of a hepatitis B virus genome of subtype *adw* isolated from a Philippino: Comparison with the reported three genomes of the same subtype. *J. Gastro. Hepatology*; **3:** 215-222.

Fattovich, G., McIntyre, G., Thursz, M., Colman, K., Giuliano, G., Alberti, A., Thomas, H. C. and Carman, W.F. (1995). Hepatitis B virus precore/core variation and interferon therapy. *Hepatology;* **22:** 1355-1362.

Feitelson, M.A. and Duan, L.-X. (1997). Hepatitis B virus X antigen in the pathogenesis of chronic infections and the development of hepatocellular carcinoma. *Am. J. Pathol.;* **150:** 1141-1157.

Felsenstein, J. (1993). PHYLIP: phylogeny inference package (Version 3.5c). Distributed by the author. Department of Genetics, University of Washington, Seattle, Washington, U.S.A.

Fernholz, D., Wildner, G., and Will, H. (1993). Minor envelope proteins of duck hepatitis B virus are initiated at internal preS AUG codons but are not essential for infectivity. *Virology;* **197:** 64-73.

Ferrari, C., Mondelli, M.U., Penna, A., Fiaccadori, F. and Chisari, F.V. (1987). Functional characterization of cloned intrahepatic, hepatitis B virus nucleoprotein-specific helper T-cell lines. *J Immunol;* **139:** 539-544.

Ferrari, C., Chisari, F.V., Ribera, E., Penna, A. and Mondelli, M.U. (1988). Functional modulation of hepatitis B core antigen-specific T lymphocytes by an autoreactive T-cell clone. *J. Immunol.*; **141:** 1155-1160.

Ferrari, C., Penna, A., Bertoletti, A., Valli, A., Antoni, A.D., Giuberti, T., Cavalli, A., Petit, M.-A. and Fiaccadori, F. (1990). Cellular immune response to hepatitis B virus-encoded antigens in acute and chronic hepatitis B virus infection. *J. Immunol.;* **145:** 3442-3449.

Ferrari, C., Bertoletti, A., Penna, A., Cavalli, A., Valli, A., Missale, G., Pilli, M., Fowler, P., Giuberti, T., Chisari, F.V. and Fiaccadori, F. (1991). Identification of immunodominant T-cell epitopes of the hepatitis B virus nucleocapsid antigen. *J. Clin. Invest.;* **88:** 214-222.

Ferrari, C., Cavalli, A., Penna, A., Valli, A., Bertoletti, A., Pedretti, G., Pilli, M., Vitali, P., Neri, T.M., Giuberti, T. and Fiaccadori, F. (1992). Fine specificity of the human T-cell response to the hepatitis B virus preS1 antigen. *Gastroenterology;* **103:** 255-263.

Foster, G.R., Ackrill, A.M., Goldin, R.D., Kerr, I.M., Thomas, H.C. and Stark, G.R. (1991).

Expression of the terminal protein region of hepatitis B virus inhibits cellular responses to interferons α and γ and double-stranded RNA. *Proc. Natl. Acad. Sci.;* **88:** 2888-2892.

Foster G.R. and Thomas H.C. (1993). Recent advances in the molecular biology of hepatitis B virus: Mutant virus and the host response. *Gut*; **34:** 1-3.

Fourel, G., Trepo, C.,Bougueleret, L., Henglein, B., Ponzetto, A., Tiollais, P. and Buendia, M.-A. (1990). Frequent activation of N-*myc* genes by hepadnavirus insertion in woodchuck liver tumours. *Nature;* **347:** 294-298.

Franco, A., Paroli, M., Testa, U., Benvenuto, R., Peschle, C., Balsano, F. and Barnaba, V. (1992). Transferrin receptor mediates uptake and presentation of hepatitis B envelope antigen by T lymphocytes. *J. Exp. Med.;* **175:** 1195-1205.

Fujii, H., Moriyama, K., Sakamoto, N., Kondo, T., Yasuda, K., Hiraizumi, Y., Yamazaki, M., Sakaki, Y., Okochi, K. and Nakjima, E. (1992). Gly145 to Arg substitution in HBsAg of immune escape mutant of hepatitis B virus. *Biochem. Biophys. Res. Comm.;* **184:** 1152-1157.

Fukai, K., Takada, S., Yokosuka, O., Saisho, H., Omata, M. and Koike, K. (1997). Characterization of a specific region in the hepatitis B virus Enhancer I for the efficient expression of X Gene in the hepatic cell. *Virology*; **236:** 279-287.

Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P. and Charnay, P. (1979). Nucleotide sequence of the hepatitis B virus genome (subtype *ayw*) cloned in *E. coli. Nature*; **281:** 646-650.

Ganem, D. (1991). Assembly of hepadnaviral virions and subviral particles, *In Hepadnaviruses, Current Topics in Microbiology and Immunology*, vol 168 pp. 61-83. (Eds.: Mason, W.S. and Seeger, C.), Berlin: Springer-Verlag.

Georgi-Geisberger, P., Berns, Ho, Locarevic, I.F., Yu, Z.-Y., Tang, Z.-Y., Zentgraf, H. and

Schroder, C.H. (1992). Mutations on free and integrated hepatitis B virus DNA in a hepatocellular carcinoma: footprints of homologous recombination. *Oncology;* **49:** 386-395.

Gerin, J.L. (1992). Treatment of viral hepatitis. *Curr. Opin. in Infect. Dis.*; **5:** 806-810.

Gerlich, W.H. and Robinson, W.S. (1980). Hepatitis B virus contains protein attached to the 5' terminus of its complete DNA strand. *Cell*; **21:** 801-809.

Gerlich, W.H., Lu, X. and Heermann, K.H. (1993). Studies on the attachment and penetration of hepatitis B virus. *J. Hepatology;* **17:** S10-S14.

Gerner, P.R., Friedt, M., Oettinger, R., Lausch, E. and Wirth, S. (1998). The hepatitis B virus seroconversion to anti-HBe is frequently associated with HBV genotype changes and selection of preS2-defective particles in chronically infected children. *Virology;* **245:** 163-172.

Giani, G., Chiaramonte, M., Pasini, C.V., Fagiolo, U. and Naccarato, R. (1979). Hepatitis B surface antigenemia and HLA antigens. *N. Engl. J. Med.*; **300:** 1056.

Gilman, M. (1993). Pleiotropy and henchman X. *Nature;* **361:** 687-688.

Girones, R. and Miller, R.L. (1989). Mutation rate of the hepadnavirus genome. *Virology;* **170:** 595-597.

Guidotti, L.G., Martinez, V., Loh, Y.-T., Rogler, C.E. and Chisari, F.V. (1994). Hepatitis B virus nucleocapsid particles do not cross the hepatocyte nuclear membrane in transgenic mice. *J. Virol.;* **68:** 5469-5475.

Gunther, S., Li, B.-C., Miska, S., Kruger, D.H., Meisel, H. and Will, H. (1995). A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. *J. Virol.*; **69:** 5437-5444.

Gunther, S., Sommer, G., Iwanska, A. and Will, H. (1997). Heterogeneity and common features of defective hepatitis B virus genomes derived from spliced pregenomic RNA. *Virology;* **238:** 363-371.

Gunther, S., Fischer, L., Pult, I., Sterneck, M. and Will, H. (1999). Naturally occurring variants of hepatitis B virus. In: *Advances in viral research.* 1st ed. Vol. 52 pp 25-137. (Eds: Maramorosch, K., Shatkin, A.J. and Murphy, F.A) San Diego: Academic Press.

Hahn, B.H., Robertson, D.L. and Sharp, P.M. (1995). Intersubtype recombination in HIV-1 and HIV-2. In *Human Retro. Aids*, pp. III-22 - III-29. (Eds: Myers, G., Hahn, B.H., Mellors, J.W., Henderson, L.E., Korber, B., Jeang, K-T., McCutchan, F.E. and Pavlakis, G.N.) New Mexico: Los Alamos National Laboratory.

Hammond, M.G., du Toit, E.D., Sanchez-Mazas, A., Andrien, M., Coluzzi, M., de Pablo, M. R., de Stefano, G., Kaplan, C., Kennedy, L.J., Louie, L. and Migot, F. (1997). ***HLA in sub-Saharan Africa.*** 12th International Histocompatibility World Conference SSAF report.

Hannoun, C., Norder, H. and Lindh, M. (2000). An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam. *J. Gen. Virol.*; **81:** 2267-2272.

Hardie, D.R., Kannemeyer, J. and Stannard, L.M. (1996). DNA single strand conformation polymorphism identifies five defined strains of hepatitis B virus (HBV) during an outbreak of HBV infection in an oncology unit. *J. Med. Virol.*; **49:** 49-54.

Hardie, D.R. and Williamson, C. (1997). Analysis of the preS1 gene of hepatitis B virus (HBV) to define epidemiologically linked and un-linked infections in South Africa. *Arch. Virol.*; **142:** 1829-1841.

Harrison, T.J., Oon, C.-J. and Zuckerman, A.R. (1991). Independent emergence of a vaccine-induced escape mutant of hepatitis B virus. *J. Hepatology;* **13:** S105-S107.

Heermann, K.H., Goldmann, U., Schwartz, W., Seyffarth, T., Horst, B. and Gerlich, W.H.

(1984). Large surface proteins of hepatitis B virus containing the Pre-s sequence. *J. Virol;* **52:** 396-402.

Heijtink, R.A., van Hattum, J., Schalm, S.W. and Masurel, N. (1982). Co-occurrence of HBsAg and anti-HBs: Two consecutive infections or a sign of advanced chronic liver disease? *J. Med. Virol.*; **10:** 83-90.

Hertogs, K., Leenders, W.P.J., Depla, E., De Bruin W.C., Meheus, L., Raymackers, J., Moshage, H. and Yap, S.H. (1993). Endonexin II, present on human liver plasma membranes, is a specific binding protein of small hepatitis B virus (HBV) envelope protein. *Virology*; **197:** 549-557.

Hildt, E., Saher, G., Bruss, V. and Hofschneider, P.H. (1996). The hepatitis B virus large surface protein (LHBS) is a transcriptional activator. *Virology;* **225:** 235-239.

Hilleman, M.R. (1994). Comparative Biology and pathogenesis of AIDS and hepatitis B viruses: Related but different. *AIDS Res. Hum. Retroviruses.*; **10:** 1409-1419.

Hino, O., Tabata, S. and Hotta, Y. (1991). Evidence for increased *in vitro* recombination with insertion of human hepatitis B virus DNA. *Proc. Natl. Acad. Sci. USA*; **88:** 9248-9252.

HLA 1991. 1st ed. Vol. 1. (Eds: Tsuji, K., Aizawa, M. and Sasazuki, T.) Oxford: Oxford University Press.

HLA 1998. 1st ed. Vol. 1. (Eds: Gjertson, David W. and Terasaki, Paul I.) Lenexa, Kansas: American Society for Histocompatibility and Immunogenetics Press.

Hodges, M., Sanders, E. and Aitken, C. (1998). Seroprevalence of hepatitis markers; HAV, HBV, HCV and HEV amongst primary school children in Freetown, Sierra Leone. *West Afr. J. Med.;* **17:** 36-37.

Hohler, T., Gerken, G., Notghi, A., Lubjuhn, R., Taheri, H., Protzer, U., Lohr, H.F., Schneider, P. M., Meyer zum Büschenfelde, K.-H. and Rittner, C. (1997). HLA-DRB1*1301 and *1302 protect against chronic hepatitis B. *J. Hepatology*; **26:** 503-507.

Hohne M., Schaefer, S., Seifer, M., Feitelson, M.A., Paul, D. and Gerlich, W.H. (1990). Malignant transformation of immortalized transgenic hepatocytes after transfection with hepatitis B virus DNA. *EMBO J.;* **9:** 1137-1145.

Hollinger, F.B. (1990). Hepatitis B virus. In: *Virology*. 2nd ed. pp. 2171-2236. (Eds: Fields, B.N. and Knipe, D.M) New York: Raven Press, Ltd.

Hosono, S., Tai, P.-C., Wang, W., Ambrose, M., Hwang Guang-Yu, D., Yuan, T.-T., Peng, B.-H., Yang, C.-S., Lee, C.-S. and Shih, C. (1995). Core antigen mutations of human hepatitis B virus in hepatomas accumulate in MHC class II-restricted T-cell epitopes. *Virology*; **212:** 151-162.

Hruska, J.E., Clayton, D.A., Rubenstein, J.L.R. and Robinson, W.S. (1977). Structure of hepatitis B Dane particle DNA before and after the Dane particle polymerase reaction. *J. Virol.*; **21:** 666-672.

Hsu, L.-C., Lin, S.-R., Hsu, H.-M., Chao, W.-H., Hsieh, J.-T., Wang, M.-C., Lu, C.-F., Chang, Y.-H. and Ho, M.-S. (1996). Ethnic differences in immune responses to hepatitis B vaccine. *Am. J. Epid.*; **143:** 718-724.

http://www.africasouth.co.za

http://encarta.msn.com

http://histo.chu-stlouis.fr

http://www.lib.ohio-state.edu

http://www.ukans.edu

Hu, J. and Seeger, C. (1996). Hsp90 is required for the activity of a hepatitis B virus reverse

transcriptase. *Proc. Natl. Acad. Sci. USA;* **93:** 1060-1064.

Hu, K.-Q., Vierling, J.M. and Siddiqui, A. (1990). *Trans*-activation of HLA-DR gene by hepatitis B virus X gene product. *Proc. Natl. Acad. Sci. USA*; **87:** 7140-7144.

Hu, X., Margolis, H.S., Purcell, R.H., Ebert, J. and Robertson, B.H. (2000). Identification of hepatitis B virus indigenous to chimpanzees. *Proc. Natl. Acad. Sci. USA;* **97:** 1661-1664.

Huang, M. and Summers, J. (1994). *pet*, a small sequence distal to the pregenome cap site, is required for expression of the duck hepatitis B virus pregenome. *J. Virol.*; **68:** 1564-1572.

Hunt, C.M., McGill, J.M., Allen, M.I. And Condreay, L.D. (2000). Clinical relevance of hepatitis B viral mutations. *Hepatology;* **31:** 1037-1044.

Hyams, K.C., al-Arabi, M.A., al-Tagani, A.A., Messiter, J.F., al-Gaali, A.A. and George, J. F. (1989). Epidemiology of hepatitis B in the Gezira region of Sudan. *Am. J. Trop. Med. Hyg.;* **40:** 200-206.

Ibara, J.R., Itoua-Ngaporo, A., Maniez-Montreuil, M., Dupressoir, M.V., Huart, J.J. and Paris, J.C. (1991). Serological markers of hepatitis B virus in hospitalized patients. *Med. Trop.;* **51:** 9-13.

Imai, M., Yanase, Y., Nojiri, T., Miyakawa, Y. and Mayumi, M.A. (1979). A receptor for polymerized human and chimpanzee albumin on hepatitis B virus particles co-occuring with HBeAg. *Gastroenterology;* **76:** 242-247.

Imanishi, T.; Wakisaka, A. and Gojobori, T. (1992). Genetic relationships among various human populations indicated by MHC polymorphisms. In: *HLA 1991*. 1st ed. Vol. I pp. 627-632. (Eds: Tsuji, K., Aizawa, M. and Sasazuki, T.). Oxford: Oxford University Press.

Jenkins, T., Zoutendyk, A. and Steinberg, A.G. (1970). Gammaglobulin groups (Gm and Inv) of various southern African populations. *Am. J. Phys. Anthrop.;***32:** 197-218.

Jin, Y., Shih, J.W.-K. and Berkower, I. (1988). Human T-cell response to the surface antigen of HBV (HBsAg). *J. Exp. Med.;* **168:** 293-306.

Jung, M.-C., Diepolder, H.M., Spengler, U., Weirenga, E.A., Zachoval, R., Hoffmann, R.M., Eichenlaub, D., Frosner, G., Will, H. and Pape, G.R. (1995). Activation of a heterogenous hepatitis B (HB) core and e antigen-specific CD4+ T-cell population during seroconversion to anti-HBe and Anti-HBs in hepatitis B virus infection. *J. Virol.*; **69:** 3358-3368.

Kaneko, S. and Miller, R.H. (1988). X-region-specific transcript in mammalian hepatitis B virus-infected liver. *J. Virol.*; **62:** 3979-3984.

Kao, J.H., Chen, P.J., Lai, M.Y. and Chen, D.S. (2000). Hepatitis B genotypes correlate with clinical outcomes in patients with chronic hepatitis B. *Gastroenterology*; **118:** 554-559.

Kao, J.H., Liu, C.—J. and Chen, D.-S. (2002). Hepatitis B viral genotypes and lamivudine resistance. *Hepatology;* **36:** 303-305.

Katkov, W.N., Watkins, E., DeMelia, H.C. and Dienstag, J.L. (1994). Immunogenicity of a "pre-S2 plus S" hepatitis B vaccine in healthy adults. *J. Viral Hepat.*; **1:** 79-83.

Kato, H., Orito, E., Sugauchi, F., Ueda, R., Gish, R.G., Usuda, S., Miyakawa, Y. and Mizokami, M. (2001). Determination of hepatitis B virus genotype G by polymerase chain reaction with hemi-nested primers. *J. Virol. Methods*; **98:** 153-159.

Kay, A., Mandart, E., Trepo, C. and Galibert, F. (1985). The HBV HBx gene expressed in E.coli is recognised by sera from hepatitis patients. *EMBO;* **4:** 1287-1292.

Kekule, A.S., Lauer, U., Weiss, L., Hofschneider, P.H. and Koshy, R. (1992). *Trans*-activation by hepatitis B virus X protein is mediated via a tumour promoter pathway. *Arch. Virol.;* **Suppl. 4:** 63-64.

Kekule, A.S., Lauer, U., Weiss, L., Luber, B. and Hofschneider, P.H. (1993). Hepatitis B virus *trans*-activator HBx uses a tumour promoter signalling pathway. *Nature;* **361:** 742-745.

Kew, M.C., Miller, R.H., Chen, H.-S., Tennant, B.C. and Purcell, R.H. (1993). Mutant woodchuck hepatitis virus genomes from virions resemble rearranged hepadnaviral integrants in hepatocellular carcinoma. *Proc. Natl. Acad. Sci. USA;* **90:** 10211-10215.

Kew, M.C. (1998). Hepatitis viruses and hepatocellular carcinoma. *Res. Virol.;* **149:** 257-262.

Kidd-Ljunggren, K., Courouce, A.-M., Oberg, M. and Kidd, A.H. (1994). Genetic conservation within subtypes in the hepatitis B virus pre-S2 region. *J. Gen. Virol.*; **75:** 1485-1490.

Kidd-Ljunggren, K., Oberg, M. and Kidd, A.H. (1995). The hepatitis B virus X gene: analysis of functional domain variation and gene phylogeny using multiple sequences. *J. Gen. Virol.*; **76:** 2119-2130.

Kiire, C.F. (1990). African Regional Study: hepatitis B infection in sub-Saharan Africa. *Vaccine*; 8: S107-S112.

Kiire, C.F. (1993). The Epidemiology and Control of hepatitis B in Sub-Saharan Africa. In: *Progress in Medical Virology* Vol. 40 pp. 141-156. (Ed: Melnick, J.L.) Basel: Karger.

Kim, C.M., Koike, K., Saito, I, Miyamura, T. and Jay, G. (1991). HBx gene of hepatitis B virus induces liver cancer in transgenic mice. *Nature;* **351:** 317-320.

Kim, S.O., Park, J.G. and Lee, Y.I. (1994). Increased expression of the insulin-like growth factor I (IGF-I) receptor gene in hepatocellular carcinoma cell lines: implications of IGF-I receptor gene activation by hepatitis B virus X gene product. *Cancer Res.*, **56:** 3831-3836.

Kock, J. and Schlicht, H. (1993). Analysis of the earliest steps of hepadnavirus replication: Genome repair after infectious entry into hepatocytes does not depend on viral polymerase activity. *J. Virol.*;**67:** 4867-4874.

Koike, K., Shirakata, Y., Yaginuma,K., Arii, M., Takada, S., Nakamura, I., Hayashi, Y., Kawada, M. and Kobayashi, M. (1989). Oncogenic potential of hepatitis B virus. *Mol. Biol. Med.*; **6:** 151-160.

Koike K. (1995). Hepatitis B virus HBx gene and hepatocarcinogenesis. *Intervirology*; **38:** 134-142.

Koschel, M., Thomssen, R. and Bruss, V. (1999). Extensive mutagenesis of the hepatitis B virus core gene and mapping of mutations that allow capsid formation. *J. Virol.*; **73:** 2153-2160.

Koshy, R. and Meyer, M. (1992). Oncogenicity of hepatitis B virus. *Rev. Med. Virol.*; **2:** 132-140.

Kramvis, A., Bukofzer, S., Kew, M.C. and Song, E. (1997). Nucleic acid sequence analysis of the Precore region of hepatitis B virus from sera of southern African black adult carriers of the virus. *Hepatology;* **25:** 235-240.

Kremsdorf, D., Garreau, F., Capel, F., Petit, M.-A. and Brechot, C. (1996). *In vivo* selection of a hepatitis B virus mutant with abnormal viral protein expression. *J. Gen. Virol.*; **77:** 929-939.

Kuroki K., Russnak R. and Ganem, D. (1989). Novel N-terminal amino acid sequence required for retention of a hepatitis B virus glycoprotein in the endoplasmic reticulum. *Mol. Cell Biol.*, **9:** 4459-4466.

Lai, M.E., Melis, A., Mazzoleni, A.P., Farci, P. and Balestrieri, A. (1991). Sequence analysis of hepatitis B virus genome of a new mutant of *ayw* subtype isolated in Sardinia. *Nucleic Acids Res.*; **19:** 5078.

Landers, T.A., Greenberg, H.B. and Robinson, W.S. (1977).Structure of the hepatitis B

Dane particle DNA and nature of the endogenous DNA polymerase reaction. *J. Virol.*; **23:** 368-376.

Lanford, R.E., Notvall, L., Lee, H. and Beames, B. (1997). *Trans*-complementation of nucleotide priming and reverse transcription between independently expressed TP and RT domains of the hepatitis B virus reverse transcriptase. *J. Virol.*; **71:** 2996-3004.

Lanford, R.E., Chavez, D., Brasky, K.M., Burns, R.B. and Rico-Hesse, R. (1998). Isolation of a hepadnavirus from the woolly monkey, a New World primate. *Proc. Natl. Acad. Sci. USA;* **95:** 5757-5761.

Lanford, R.E., Kim, Y.H., Lee, H., Notvall, L. and Beames, B. (1999). Mapping of the hepatitis B virus reserves transcriptase TP and RT domains by transcomplementation for nucleotide priming and by protein-protein interaction. *J. Virol.*; **73:** 1885-1893.

Lanford, R.E., Chavez, D., Rico-Hesse, R. and Mootnick, A. (2000). Hepadnavirus infection in captive gibbons. *J. Virol.;* **74:** 2955-2959.

Lango-Warensjo, A., Cardell, K. and Lindblom, B. (1998). Haplotypes comprising subtypes of the DQB1*06 allele direct the antibody response after immunisation with hepatitis B surface antigen. *Tissue Antigens*; **52:** 374-380.

Lara-Pezzi, E., Majano, P.L., Gomez-Gonzalo, M., Garcia-Monzon, C., Moreno-Otero, R., Levero, M. and Lopez-Cabrera, M. (1998). The hepatitis B virus X protein up-regulates tumor necrosis factor alpha gene expression in hepatocytes. *Hepatology;* **28:** 1013-1021.

Lau, J.Y.N. and Wright, T.L. (1993). Molecular virology and pathogenesis of hepatitis B. *The Lancet*; **342:** 1335-1339.

Lauder, I.J., Lin, H-J., Lau, J.Y.N., Siu, T-S. and Lai, C-L. (1993). The variability of the hepatitis B virus genome: statistical analysis and biological implications. *Mol. Biol. Evol.;* **10:** 457-470.

Lee, T.H., Elledge, S.J. and Butel, J.S. (1995). Hepatitis B virus X protein interacts with a probable cellular DNA repair protein. *J. Virol.;* **69:** 1107-1114.

Lenhoff, R.J. and Summers, J. (1994). Construction of avian hepadnavirus variants with enhanced replication and cytopathicity in primary hepatocytes. *J. Virol.;* **68:** 5706-5713.

Li, J.-S., Tong, S.-P., Wen, Y.-M., Vitvitski, L., Zhang, Q. and Trepo, C. (1993). Hepatitis B virus genotype A rarely circulates as an HBe-minus mutant: possible contribution of a single nucleotide in the preCore region. *J. Virol.*; **67:** 5402-5410

Lin, Y., Nomura, T., Cheong, J., Dorjsuren, D., Iida, K. and Murakami, S. (1997). Hepatitis B virus X protein is a transcriptional modulator that communicates with transcription factor IIB and the RNA polymerase II subunit 5*. *J. Biol. Chem.;* **272:** 7132-7139.

Lindh, M., Andersson, A.-S. and Gusdal, A. (1997). Genotypes, nt 1858 variants and geographic origin of hepatitis B virus - large-scale analysis using a new genotyping method. *J. Infect. Dis.*; **175:** 1285-1293.

Lindh, M., Gonzalez, J.E., Norkrans, G. and Horal, P. (1998). Genotyping of hepatitis B virus by restriction pattern analysis of a preS amplicon. *J. Virol. Meth.*; **72:** 163-174.

Lo, B.B., Meymouna, M., Boulahi, M.A., Tew, M., Sow, A., Ba, A. and Sow, M.B. (1999). Prevalence of serum markers of hepatitis B and C virus in blood donors of Nouakchott, Mauritania. *Bull. Soc. Pathol. Exot.;* **92:** 83-84.

Lohoues-Kouacou, M.J., Toure, M., Hillah, J., Camara, B.M., N'Dri, N., Kouame, K.J. and Attia, Y. (1998). Materno-fetal transmission of hepatitis B virus in Ivory Coast. Plea for mass vaccination. *Sante;* **8:** 401-404.

Lohr, H., Weber, W., Schlaak, J., Goergen, B., Meyer zum Buschenfelde, K.-H. and Gerken, G. (1995). Proliferative response of CD4+ T-cells and hepatitis B virus clearance in chronic

hepatitis with or without hepatitis B e-minus hepatitis B virus mutants. *Hepatology;* **22:** 61-68.

Luber, B., Lauer, U., Weiss, L., Hohne, M.., Hofschneider, P.H. and Kekule, A.S. (1993). The hepatitis B virus transactivator HBx causes elevation of diacylglycerol and activation of protein kinase C. *Res. Virol.;* **144:** 311-321.

MacDonald, D.M., Holmes, E.C. Lewis, J.C.M. and Simmonds, P. (2000). Detection of hepatitis B virus infection in wild-born chimpanzees (*Pan troglodytes verus)*: Phylogenetic relationships with human and other primate genotype. *J. Virol.;* **74:** 4253-4257.

Magnius, L.O. and Norder, H. (1995). Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variablility of the S-gene. *Intervirology*; **38:** 24-34.

Maguire, H.F., Hoeffer, J.P. and Siddiqui, A. (1991). HBV X protein alters the DNA binding specificity of CREB and ATF-2 by protein-protein interactions. *Science;* **252:** 842-844.

Mahe, Y., Mukaida, N., Kuno, K., Akiyama, M., Ikeda, N., Matsushima, K. and Murakami, S. (1991). Hepatitis B virus X protein *trans*-activates human interleukin-8 gene through acting on nuclear factor κB and CCAAT/enhancer-binding protein-like *cis*-elements. *J. Biol. Chem.;* **25:** 13759-13763.

Mangold, C.M.T., Unckell, F., Werr, M. and Streek, R.E. (1995). Secretion and antigenicity of hepatitis B virus small envelope proteins lacking cysteines in the major antigenic region. *Virology;* **211:** 535-543.

Mangold, C.M.T., Unckell, F., Werr, M. and Streek, R.E. (1997). Analysis of intermolecular disulfide bonds and free sulfhydryl groups in hepatitis B surface antigen particles. *Arch. Virol.:* **142:** 2257-2267.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989). *Cloning: A Laboratory Manual,* 2nd Edition. New York: Cold Spring Harbor Laboratory Press.

Margalit, H., Spouge, J.L., Cornette, J.L., Cease, K.B., Delisi, C. and Berzofsky, J.A. (1987). Prediction of immunodominant helper T-cell antigenic sites from the primary sequence. *J. Immunol.*; **138:** 2213-2229.

Marion, P.L., Van Davelaar, M.J., Knight, S.S., Salazar, F.H., Garcia, G., Popper, H. and Robinson, W.S. (1986). Hepatocellular carcinoma in ground squirrels persistently infected with ground squirrel hepatitis virus. *Proc. Natl. Acad. Sci. USA;* **83:** 4543-4546.

Martinson, F.E., Weigle, K.A., Royce, R.A., Weber, D.J., Suchindran, C.M. and Lemon, S. M. (1998). Risk factors for horizontal transmission of hepatitis B virus in a rural district in Ghana. *Am. J. Epid.;* **147:** 478-487.

Maruyama, T., McLachlan, A., Iino, S., Koike, K., Kurokawa, K. and Milich, D.R. (1993). The serology of chronic hepatitis B infection revisited. *J. Clin. Invest.;* **91:** 2586-2595.

Mason, W.S., Halpern, M.S., England, J.M., Seal, G., Egan, J., Coates, L., Aldrich, C. and Summers, J. (1983). Experimental transmission of duck hepatitis B virus. *Virology*; **131:** 375-384.

Matee, M.I., Lyamuya, E.F., Mbena, E.C., Magessa, P.M., Sufi, J., Marwa, G.J., Mwasulama, O.J. and Mbwana, J. (1999). Prevalence of transfusion-associated viral infections and syphilis among blood donors in Muhimbili Medical Centre, Dar es Salaam, Tanzania. *East Afr. Med. J.*; **76:** 167-171.

Maynard, J.E. (1990). Hepatitis B: global importance and need for control. *Vaccine;* **Suppl. 8:** S18-S20.

Mbayed, V.A., Lopez, J.L., Telenta, P.F.S., Palacios, G., Badía, I., Ferro, A., Galoppo, C. and Campos, R. (1998). Distribution of hepatitis B virus genotypes in two different pediatric

populations from Argentina. *J. Clin. Micro.*; **36:** 3362-3365.

Mehdi, H., Kaplan, M.J., Anlar, F.Y., Yang, X., Bayer, R., Sutherland, K. and Peeples, M.E. (1994). Hepatitis B virus surface antigen binds to apolipoprotein H. *J. Virol.*; **68:** 2415-2424.

Mehdi, H., Yang, X. and Peeples, M.E. (1996). An altered form of apolipoprotein H binds hepatitis B virus surface antigen most efficiently. *Virology*; **217:** 58-66.

Meisel, H., Sominskaya, I., Pumpen, P., Pushko, P., Borisove, G., Deepen, R., Lu, P., Spiller, G.H., Kruger, D.H., Grens, E. and Gerlich, W.H. (1994). Fine mapping and functional characterization of two immuno dominant regions from the preS2 sequence of hepatitis B virus. *Intervirology*; **37:** 330-339.

Meister, G.E., Roberts, C.G.P., Berzofsky, J.A. and De Groot, A.S. (1995). Two novel T-cell epitope prediction algorithms based on MHC-binding motifs; comparison of predicted and published epitopes from *Mycobacterium tuberculosis* and HIV protein sequences. *Vaccine*; **13:** 581-591.

Melegari, M., Scaglioni, P.P. and Wands, J.R. (1997). The small envelope protein is required for secretion of a naturally occurring hepatitis B virus mutant with preS1 deleted. *J. Virol.;* **71:** 5449-5454.

Meuer, S.C. and Moebius, U. (1994). Immune mechanisms underlying HBV persistence. *Semin. Virol.*; **5:** 289-295.

Meunier, D.M., Vohito, M.D., Courrouce, A.M., Lesbordes, J.L., Georges-Courbot, M.C. and Georges, A.J. (1984). Prevalence of hepatitis B surface antigen and determination of its subtypes in the Central African Republic. *Bull. Soc. Pathol. Exot. Filiales*; **77:** 249-254.

Michalak, T.I., Pasquinelli, C., Guilhot, S. and Chisari, F.V. (1994). Hepatitis B virus persistence after recovery from acute viral hepatitis. *J. Clin. Invest.*;**93:** 230-239.

Milich, D.R., McLachlan, A., Thornton, G.B. and Hughes, J.L. (1987). Antibody production to the nucleocapsid and envelope of the hepatitis B virus primed by a single synthetic T-cell site. *Nature*; **329:** 547-549.

Milich, D.R., Jones, J.E., Hughes, J.L., Price, J., Raney, A.K. and McLachlan, A. (1990). Is a function of the secreted hepatitis B e antigen to induce immunologic tolerance *in utero*. *Proc. Natl. Acad. Sci. USA;* **87:** 6599-6603.

Milich, D.R., Jones, J., Hughes, J. and Maruyama, T. (1993). Role of T-cell tolerance in the persistence of hepatitis B virus infection. *J. Immunotherapy*; **14:** 226-233.

Milich, D.R., Sallberg, M. and Maruyama, T. (1995). The humoral immune response in acute and chronic hepatitis B virus infection. *Semin. Immunopathol.;* **17:** 149-166.

Milich, D.R., Chen, M., Schodel, F., Peterson, D.L., Jones, J.E. and Hughes, J.L. (1997). Role of B cells in antigen presentation of the hepatitis B core. *Proc. Natl. Acad. Sci. USA;* **94:** 14648-14653.

Miller, R.H. and Robinson, W.S. (1986). Common Evolutionary Origin of hepatitis B virus and retroviruses. *Proc. Soc. Exp. Biol. Med.*; **83:** 2531-2535.

Miller, R.H. (1988). Close evolutionary relatedness of the hepatitis B virus and murine leukemia virus polymerase gene sequences. *Virology*; **164:** 147-155.

Missale, G., Redeker, A., Person, J., Fowler, P., Guilhot, S., Schlicht, H.J., Ferrari, C. and Chisari, F.V. (1993). HLA-A31- and HLA-Aw68-restricted cytotoxic T-cell responses to a single hepatitis B virus nucleocapsid epitope during acute viral hepatitis. *J. Exp. Med.;* **177:** 751-762.

Mizokami, M., Orito, E., Ohba, K.-I., Ikeo, K., Lau, J.Y.N. and Gojobori, T. (1997). Constrained evolution with respect to gene overlap of hepatitis B virus. *J. Mol. Evol.*; **44:** S83-S90.

Mizokami, M., Nakano, T., Orito, E., Tanaka, Y., Sakugawa, H., Mukaide, M. and Robertson, B. H. (1999). Hepatitis B virus genotype assignment using restriction fragment length polymorphism patterns. *FEBS Lett.*; **450:** 66-71.

Moraes, M.T.B., Gomes, S.A. and Niel, C. (1996). Sequence analysis of preS/S gene of hepatitis B virus strains of genotypes A, D and F isolated in Brazil. *Arch. Virol.*; **141:** 1767-1773.

Moriarty, A.M., Alexander, H., Lerner, R.A. and Thornton, G.B. (1985). Antibodies to peptides detect new hepatitis B antigen serological correlation with hepatocellular carcinoma. *Science;* **227:** 429-432.

Moroy, T., Entiemble, J., Trepo, C., Tiollais, P. and Buendia, M.A. (1985). Transcription of woodchuck hepatitis virus in the chronically infected liver. *EMBO;* **4:** 1507-1514.

Mota, A.H., Fainboim, H., Terg, R. and Fainboim, L. (1987). Fifty-one patients with chronic active hepatitis were typed for their HLA-A, B, C and DR antigens. We observed a significant increase in the antigen frequency of HLA B35 in patients compared with controls. *Tissue Antigens*; **30:** 238-240.

Nassal, M., Junker-Niepmann, M. and Schaller, H. (1990). Translational inactivation of RNA function: Discrimination against a subset of genomic transcripts during HBV nucleocapsid assembly. *Cell.*; **63:** 1357-1363.

Nassal, M. and Schaller, H. (1996). Hepatitis B virus replication - an update. *J. Viral Hepat.*; **3:** 217-226.

Nayersina, R., Fowler, P., Guilhot, S., Missale, G., Cerny, A., Schlicht, H.-J., Vitiello, A., Chesnut, R., Person, J.L., Redeker, A.G., and Chisari, F.V. (1993). HLA A2 restricted cytotoxic T lymphocyte responses to multiple hepatitis B surface antigen epitopes during hepatitis B virus infection. *J. Immunol.;* **150:** 4659-4671.

Neurath, A.R., Strick, N. and Sproul, P. (1992). Search for hepatitis B virus cell receptors reveals binding sites for interleukin 6 on the virus envelope protein. *J. Exp. Med.;* **175:** 461-469.

Niel, C., Moraes, M.T.B., Gaspar, A.M.C., Yoshida, C.F.T. and Gomes, S.A. (1994). Genetic diversity of hepatitis B virus strains isolated in Rio de Janeiro, Brazil. *J. Med. Virol.*; **44:** 180-186.

Norder, H., Hammas, B. and Magnius, L.O. (1990). Typing of hepatitis B virus genomes by a simplified polymerase chain reaction. *J. Med. Virol.*; **31:** 215-221.

Norder, H., Hammas, B., Lofdahl, S., Courouce, A.- M. and Magnius, L.O. (1992). Comparison of the amino acid sequences of nine different serotypes of hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains. *J. Gen. Virol.*; **73:** 1201-1208.

Norder, H., Hammas, B., Shou-Dong, L., Bile, K., Courouce, A.-M., Mushahwar, I.K. and Magnius, L.O. (1993). Genetic relatedness of hepatitis B viral strains of diverse geographical origin and natural variations in the primary structure of the surface antigen. *J. Gen. Virol.*; **74:** 1341-1348.

Norder, H., Courouce, A.-M. and Magnius, L.O. (1994). Complete genomes, phylogenetic relatedness and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virology;* **198:** 489-503.

Norder, H., Ebert, J.W., Fields, H.A., Mushahwar, I.K. and Magnius, L.O. (1996). Complete sequencing of a gibbon hepatitis B virus genome reveals a unique genotype distantly related to the chimpanzee hepatitis B virus. *Virology;* **218:** 214-223.

Nurse, G.T., Bodmer, J.G., Bodmer, W.F., van Leeuwen, A., van Rood, J.J., du Toit, E.D. and Botha, M.C. (1975). A reassessment of the HLA system in Khoisan populations of South West Africa. *Tissue Antigens;* **5:** 402-414.

Oberhaus, S.M. and Newbold, J.E. (1995). Detection of an RNase H activity associated with hepadnaviruses. *J. Virol.*; **69:** 5697-5704.

Obert S., Zachmann-Brand B., Deindl E., Tucker W., Bartenschlager R. and Schaller H.A. (1996). A splice hepadnavirus RNA that is essential for virus replication. *EMBO J*; **15**: 2565-2574.

Ohba, K.-I., Mizokami, M., Ohno, T., Suzuki, K., Orito, E., Lau, J.Y.N., Ina, Y., Ikeo, K. and Gojobori, T. (1995). Relationships between serotypes and genotypes of hepatitis B virus: genetic classification of HBV by use of surface genes. *Virus Res.*; **39:** 25-34.

Okamoto, H., Imai, M., Tsuda, F., Tanaka, T., Miyakawa, Y. and Mayumi, M. (1987). Point mutation in the S-gene of hepatitis B virus for a *d/y* or *w/r* subtypic determinant in two blood donors carrying a surface antigen of compound subtype *adyr* and *adwr. J. Gen. Virol.;* **61:** 3030-3034.

Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R.I., Imai, M., Miyakawa,Y., Mayumi, M. (1988). Typing hepatitis B virus by homology in nucleotide sequence: Comparison of surface antigen subtypes. *J. Gen.Virol.*; **69:** 2575-2583.

Okamoto, H., Omi, S., Wang, Y., Itoh, Y., Tsuda, F., Tanaka, T., Akahane, Y., Miyakawa, Y. and Mayumi, M. (1989). The loss of subtypic determinants in alleles *d/y* or *w/r* on hepatitis B surface antigen. *Mol. Immunol.*; **26:** 197-205.

Okoye, R.C., Willliams, E., Alonso, A., Doyle, P., Awad, J., Navarrette, C., Jaraquemada, D., Ollier, W.E.R. and Festenstein, H. (1985). HLA polymorphisms in Nigerians. *Tissue Antigens;* **5:** 402-414.

Oliver, R. (1991). *The African Experience.* 1st ed. Chatham Kent, Great Britain: Pimlico.

Oon, C.J., Lim, G.K., Ye, Z., Goh, K.T., Tan, K.L., Yo, S.L., Hopes, E., Harrison, T.J. and Zuckerman, A.J. (1995). Molecular epidemiology of hepatitis B virus vaccine variants in Singapore. *Vaccine;* **13:** 699-702.

Orito, E., Mizokami, M., Sakugawa, H., Michitaka, K., Ishikawa, K., Ichida, T., Okanoue, T., Yotsuyanagi, H. and Iino S. (2001). A case-control study for clinical and molecular biological differences between hepatitis B viruses of genotypes B and C. ***Hepatology;*** **33:** 218-223.

Ou, J.- H., Laub, O. and Rutter, W.J. (1986). Hepatitis B virus gene function: The preCore region targets the core antigen to cellular membranes and causes the secretion of the e antigen. *Proc. Natl. Acad. Sci. USA;* **83:** 1578-1582.

Owiredu, W.K.B.A., Kramvis, A. and Kew, M.C. (2001). Hepatitis B virus DNA in serum of healthy black African adults positive for hepatitis B surface antibody alone: possible association with recombination between genotypes A and D. *J. Med. Virol.*; **64:** 441-454.

Parker, K.C., Bednarek, M.A. and Coligan, J.E. (1994). Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.;* **152:** 163-175.

Pathak, V.K. and Wei-Shau, H. (1997). "Might as well jump!" Template switching by retroviral reverse transcriptase, defective genome formation and recombination. ***Semin.*** *Virol.*; **8:** 141-150.

Patzel, V. and Sczakiel, G. (1997). The hepatitis B virus posttranscriptional regulatory element contains a highly stable RNA secondary structure. *Biochem. Biophys. Res. Commun.*; **231:** 864-867.

Penna, A., Chisari, F.V., Bertoletti, A., Missale, G., Fowler, P., Giuberti, T., Fiaccadori, F. and Ferrari, C. (1991). Cytotoxic T lymphocytes recognize an HLA-A2-restricted epitope within the hepatitis B virus nucleocapsid antigen. *J. Exp. Med.*; **174:** 1565-1570.

Petit, M.A., Capel, F., Dubanchet, S. and Mabit, H. (1992). PreS1-specific binding proteins as potential receptors for hepatitis B virus in human hepatocytes. ***Virology;*** **187:** 211-222.

Petrie, A.. (1987). Estimation. In: ***Lecture notes in medical statistics***. 2nd ed. pp. 47-55. Oxford: Blackwell Scientific Publications.

Pineau, P., Marchio, A., Terris, B., Matte, M.G., Tu, Z.X., Tiollais, P. and Dejean, A.A. (1996). t(3;8) chromosomal translocation associated with hepatitis B virus intergration involves the carboxypeptidase N locus. ***J. Virol.***; **70:** 7280-7284.

Pineau, P., Marchio, A., Mattei, M.-G., Kim, W.-H., Youn, J.-K., Tiollais, P., and Dejean, A.(1998). Extensive analysis of duplicated-inverted hepatitis B virus integrations in human hepatocellular carcinoma. ***J. Gen. Virol.;*** **79:** 591-600.

Pollicino, T., Pernice, F., Campo, S., Mesiti, O., Misefari, V., Pernice, M. and Raimondo, G. (1996). Severe outcome of hepatitis B virus (HBV) infection and lack of HBV e antigen-defective virus emergence in patients homozygous for HLA class I alleles. ***J. Gen. Virol.***; **77:** 1833-1836.

Pontisso, P., Ruvoletto, M.G., Tiribelli, C., Gerlich, W.H., Ruol, A. and Alberti, A. (1992). The preS1 domain of hepatitis B virus and IgA cross-react in their binding to the hepatocyte surface. *J. Gen. Virol.;* **73:** 2041-2045.

Population Census, 1991. Summarised results after adjustment for undercount (1991). CSS Report No. 03-01-01. Pretoria: Central Statistical Service.

Pourquier, P., Jensen, A.D., Gong, S.S., Pommier, Y. and Rogler, C.E. (1999). Human DNA topoisomerase I-mediated cleavage and recombination of duck hepatitis B virus DNA *in vitro. Nucleic Acids Res.;* **27:** 1919-1925.

Prange, R. and Streeck, R.E. (1995). Novel transmembrane topology of the hepatitis B virus envelope proteins. *EMBO*; **14:** 247-256.

Prince, A.M., White, T., Pollack, N., Riddle, J., Brotman, B. and Richardson, L. (1981).

Epidemiology of hepatitis B infection in Liberian infants. *Infect. Immun.;* **32:** 675-680.

Pult, I., Chouard, T., Wieland, S., Klemenz, R., Yaniv, M. and Blum, H.E. (1997). A hepatitis B virus mutant with a new hepatocyte nuclear factor I binding site emerging in transplant-transmitted fulminant hepatitis B. *Hepatology;* **25:** 1507-1515.

Qadri, I., Maguire, H.F. and Siddiqui, A. (1995). Hepatitis B virus *trans*-activator protein X interacts with the TATA-binding protein. *Proc. Natl. Acad. Sci. USA;* **92:** 1003-1007.

Raimondo, G., Burk, R.D., Lieberman, H.M., Muschel, J., Hadziyannis, S.J., Will, H., Kew, M.C., Dusheiko, G.M. and Shafritz, D.A. (1988). Interrupted replication of hepatitis B virus in liver tissue of HBsAg carriers with hepatocellular carcinoma. *Virology;* **166:** 103-112.

Rammensee, H.-G., Friede, T. and Stevanovic, S. (1995). MHC ligands and peptide motifs: first listing. *Immunogenetics;* **41:** 178-228.

Rammensee, H.-G., Bachmann, J., Emmerich, N. N., Bachor, O.A. and Stevanovic, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics;* **50:** 213-219.

Rehermann, B., Person, J., Redeker, A., Fowler, P., Brown, M., Sette, A., Moss, B. and Chisari, F.V. (1995a). The cytotoxic T lymphocyte response to multiple hepatitis B virus polymerase epitopes during and after acute viral hepatitis. *J. Exp. Med.;* **181:** 1047-1058.

Rehermann, B., Pasquinelli, C., Mosier, S.M. and Chisari, F.V. (1995b). Hepatitis B virus (HBV) sequence variation in cytotoxic T lymphocyte epitopes is not common in patients with chronic HBV infection. *J. Clin. Invest.;* **96:** 1527-1534.

Repp, R., Rhiel, S., Heermann, K., Schaefer, S., Keller, C., Ndumbe, P., Lampert, F. and Gerlich, W.H. (1993). Genotyping by multiplex polymerase chain reaction for detection of endemic hepatitis B virus transmission. *J. Clin. Micro.;* **31:** 1095-1102.

Richard-Lenoble, D., Traore, O., Kombile, M., Roingeard, P., Dubois, F. and Goudeau, A. (1995). Hepatitis B, C, D and E markers in rural equatorial African villages (Gabon). *Am. J. Trop. Med. Hyg.;* **53:** 338-341.

Rivkina, M.B., Lunin, V.G., Mahov, A.M., Tikchonenko, T.I. and Kukain, R.A. (1988). Nucleotide sequence of integrated hepatitis B virus DNA and human flanking regions in the genome of the PLC/PRF/5 cell line. *Gene;* **64:** 285-296.

Robinson, W.S., Clayton, D.A. and Greenman, R.L. (1974). DNA of a human hepatitis B virus candidate. *J. Virol.*; **14:** 384-391.

Rossner, M.T. (1992). Review: Hepatitis B virus X-gene product: A promiscuous transcriptional activator. *J. Med. Virol.;* **36:** 101-117.

Salfeld, J., Pfaff, E., Noah, M. and Schaller, H., (1989). Antigenic determinants and functional domains in core antigen and e antigen from hepatitis B virus. *J. Virol.*; **63:** 798-808.

Sastrosoewignjo, R.I., Sandjaja, B. and Okamoto, H. (1991). Molecular epidemiology of hepatitis B virus in Indonesia. *J. Gastro. Hepatology*; **6:** 491-498.

Scaglioni, P.P., Melegari, M. and Wands, J.R. (1997). Posttranscriptional regulation of hepatitis B virus replication by the precore protein. *J. Virol.;* **71:** 345-353.

Schaller, H. and Fischer, M. (1991). Transcriptional control of hepadnavirus gene expression. In: *Hepadnaviruses* 1st ed. Vol. 1 pp. 21-39. (Eds: Mason, WS. and Seeger, C.) Berlin: Springer-Verlag.

Schirmacher, P., Wang, W., Stahnke, G., Will, H. and Rogler, C.E. (1995). Sequences and structures at hepadnaviral integration: recombination sites implicate topoisomerase I in hepadnaviral DNA rearrangements and integration. *J. Hepatology*; **22:** 21-23.

Scolnick, E.M., McLean, A.A., West, D.J., McAleer, W.J., Miller, W.J. and Buynak, E.B. (1984). Clinical evaluation in healthy adults of a hepatitis B vaccine made by recombinant DNA. *JAMA*; **251:** 2812-2815.

Scully, L.J., Brown, D., Lloyd, C., Shein, R. and Thomas, H.C. (1990). Immunological studies before and during interferon therapy in chronic HBV infection: identification of factors predicting response. *Hepatology*; **12: 1111-1117**.

Seidel, K. (1997). *The prevalence of serological markers for hepatitis B amongst first time blood donors in Namibia. Baseline studies for immunisation.* A research report submitted in partial fulfilment of the diploma of Public Health, department of Community Health, University of the Witwatersrand, Johannesburg.

Seifer, M. and Standring, D.N. (1994). A protease-sensitive hinge linking the two domains of the hepatitis B virus core protein is exposed on the viral capsid surface. *J. Virol.*; **68:** 5548-5555.

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

Shih, J.W.- K., Cheung, L.C., Alter, H.J., Lee, L.M. and Gu, J.R. (1991). Strain analysis of hepatitis B virus on the basis of restriction endonuclease analysis of polymerase chain reaction products. *J. Clin. Micro.*; **29:** 1640-1644.

Shu, P.-Y., Lin, M.-Y., Chang, C., Hwang, L.-H., Huang, J.-H., Chang, T.-H. and Hu, C.-P. (1994). Presence of hepatitis B surface antigen (HBsAg)-specific cytotoxic T cells in asymptomatic HBsAg carriers. *J. Infect. Dis.;* **169:** 633-637.

Siepel, A.C. and Korber, B.T. (1995). Scanning the database for recombinant HIV-1 genomes. *In Human Retroviruses and Aids.* pp. III-35-III-60. (Eds: Myers, G., Hahn, B.H., Mellors, J.W., Henderson, L.E., Korber, B., Jeang, K.-T., McCutchen, F.E. and Pavlakis, G. N.). New Mexico: Los Alamos National Laboratory.

Sim, J. (1993). *Hepatitis B virus serology: Eight years of laboratory data in retrospect.* A dissertation submitted in partial fulfilment of the requirements for the degree Master of Medicine (Virology), Faculty of Medicine, University of the Witwatersrand, Johannesburg.

Simmonds, P., Mellor, J., Sakuldamrongpanich, T., Nuchaprayoon, C., Tanprasert, S., Holmes, E.C. and Smith, D.B. (1996). Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: Comparison with classifications based upon sequence similarity. *J. Gen. Virol.*; **77:** 3013-3024.

Slagle, B.L., Lee, T.H., Medina, D., Finegold, M.J. and Butel, J.S. (1996). Increased sensitivity to the hepatocarcinogen diethylnitrosamine in transgenic mice carrying the hepatitis B virus X gene. *Mol. Carcinog.*; **15:** 261-269.

Smith, J.M. (1992). Analyzing the mosaic structures of genes. *J. Mol. Evol.;* **34:** 126-129.

Snyder, R.L., Tyler, G. and Summers, J. (1982). Chronic hepatitis and hepatocellular carcinoma associated with woodchuck hepatitis virus. *Am. J. Pathol;* **107:** 422-425.

Stachowski, J., Kramer, J., Füst, G., Maciejewski, J., Baldamus, C.A. and Petrányi, G.G. (1995). Relationship between the reactivity to hepatitis B virus vaccination and the frequency of MHC class I, II and III alleles in haemodialysis patients. *Scand. J. Immunol.*; **42:** 60-65.

Standring, D.N., Ou, J.H. and Rutter, W.J. (1986). Assembly of viral particles in Xenopus oocytes: Pre-Surface-antigens regulate secretion of the hepatitis B viral surface envelope particle. *Proc. Natl. Acad. Sci. USA;* **83:** 9338-9342.

Steele, A.D. and Bos, P. (1996). Hepatitis B and C virus infection in adult volunteers in Angola. *S. Afr. Med. J.*; **86:** 701-702.

Stemler, M., Weimer, T., Tu, Z.X., Wan, D.F., Levrero, M., Jung, C., Pape, G.R. and Will,

H. (1990). Mapping of B cell epitopes of the human hepatitis B virus X protein. *J. Virol.;* **64:** 2802-2809.

Stephens, J.C. (1985). Statistical methods of DNA sequence analysis: Detection of intragenic recombination or gene conversion. *Mol. Biol. Evol.*; **2:** 539-556.

Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L. and Wiley, D.C. (1994). Crystal structrue of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature*; **368:** 215-221.

Sterneck, M., Kalinina, T., Otto, S., Günther, S., Fischer, L., Burdelski, M., Greten, H., Broelsch, C.E. and Will, H. (1998). Neonatal fulminant hepatitis B: Structural and functional analysis of complete hepatitis B virus genomes from mother and infant. *J. Infect. Dis.*; **177:** 1378-1381.

Stewart, C.-B. (1993). The powers and pitfalls of parsimony. *Nature;* **361:** 603-607.

Stirk, H.J., Thornton, J.M. and Howard, C.R. (1992). A topological model for hepatitis B surface antigen. *Intervirology*; **33:** 148-158.

Stuyver , L., De Gendt, S., Van Geyt, C., Zoulim, F., Fried, M., Schinazi ,R.F. and Rossau, R. (2000). A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *J. Gen. Virol.*; **81:** 67-74.

Su, Q., Liu, Y.-F., Zhang, J.-F., Zhang, S.-X., Li, D.-F. and Yang,J.-J. (1994). Expression of insulin-like growth factor II in hepatitis B, cirrhosis and hepatocellular carcinoma: Its relationship with hepatitis B virus antigen expression. *Hepatology;* **20:** 788-799.

Summers, J., O'Connell, A. and Millman, I. (1975). Genome of hepatitis B virus: Restriction enzyme cleavage and structure of the DNA isolated from Dane particles. *Proc. Natl. Acad. Sci. USA*; **72:** 4597-4601.

Summers, J. and Mason, W. (1982a). Properties of hepatitis B-like viruses related to their taxonomic classification. *Hepatology*; **2:** 61S-66S.

Summers, J. and Mason, W.S. (1982b). Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell;* **29:** 403-415.

Szmuness, W. (1978). Hepatocellular carcinoma and the hepatitis B virus: Evidence for a causal association. *Prog. in Med. Virol.*; **24:** 40-69.

Szmuness, W., Stevens, C.E., Harley, C.E., Zang, E.A., Oleszko, W.R., William, D.C., Sadovsky, R., Morrison, J.M. and Kellner, A. (1980). Hepatitis B vaccine: demonstration of efficacy in a controlled clinical trial in a high-risk populationin the United States. *N. Engl. J. Med.*; **303:** 833-841.

Tabor, E., Gerety, R.J., Smallwood, L.A and Barker, L.F. (1977). Coincident hepatitis B surface antigen and antibodies of different subtypes in human serum. *J. Immunol.*; **118:** 369-370.

Tabor, E., Bayley, A.C., Cairns, J., Pelleu, L. and Gerety, R.J. (1985). Horizontal transmission of hepatitis B virus among children and adults in five rural villages in Zambia. *J. Med. Virol.*; **15:** 113-120.

Takada, S. and Koike, K. (1990). *Trans*-activation function of a 3' truncated X gene-cell fusion product from integrated hepatitis B virus DNA in chronic hepatitis tissues. *Proc. Natl. Acad. Sci. U S A*; **15:** 5628-5632.

Takada, S. and Koike, K. (1994). Three sites of the hepatitis B virus X protein cooperatively interact with cellular proteins. *Virology*; **205:** 503-510.

Takahashi, K., Aoyama, K., Ohno, N., Iwata, K., Akahane, Y., Baba, K., Yoshizawa, H. and Mishiro, S. (1995). The Precore/core promoter mutant (T1762 A1764) of hepatitis B virus: clinical significance and an easy method for detection. *J Gen Virol;* **76 :** 3159-3164.

Takahashi, K., Brotman, B., Usuda, S., Mishiro, S. and Prince, A.M. (2000). Full-genome sequence analyses of hepatitis B virus (HBV) strains recovered from chimpanzees infected in the wild: Implications for an origin of HBV. *Virology*; **267:** 58-64.

Tavis, J.E., Perri, S. and Ganem, D. (1994). Hepadnavirus reverse transcription initiates within the stem-loop of the RNA packaging signal and employs a novel strand transfer. *J. Virol.*; **68:** 3536-3543.

Taylor, W.R. (1986). The classification of amino acid conservation. *J. Theor. Biol.*; **119:** 205-218.

Telenta, P.F.S., Poggio, G.P., Lopez, J.L., Gonzalez, J., Lemberg, A. and Campos, R.H. (1997). Increased prevalence of genotype F hepatitis B virus isolates in Buenos Aires, Argentina. *J. Clin. Micro.*; **35:** 1873-1875.

Terrault, N.A., Holland, C.C., Ferrell, L., Hahn, J.A., Lake, J.R., Roberts, J.P. and Ascher, N.L. (1996). Interferon alfa for recurrent hepatitis B infection after liver transplantation. *Liver Transpl. Surg.*; **2:**132-138.

Thio, C.L., Carrington, M., Marti, D., O'Brien, S.J., Vlahov, D., Nelson, K.E., Astemborski, J. and Thomas, D.L. (1999). Class II alleles and hepatitis B virus persistence in African Americans. *J. Infect. Dis.*; **179:** 1004-1006.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*; **22:** 4673-4680.

Thursz, M.R., Kwiatkowski, D., Allsopp, C.E.M., Greenwood, B.M., Thomas, H.C. and Hill, A.V.S. (1995). Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. *New Eng. J. of Med.*; **332:** 1065-1069.

Thursz, M.R. (1997a). Host genetic factors influencing the outcome of hepatitis. *J. Viral Hepat.*; **4:** 215-220.

Thursz, M.R., Thomas, H.C., Greenwood, B.M. and Hill, A.V.S. (1997b). Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nature Gen.*; **17**: 11-12.

Tiollais, P., Charnay, P. and Vyas, G.N. (1981). Biology of hepatitis B virus. *Science*; **213:** 406-411.

Tiollais, P., Pourcel, C. and Dejean, A. (1985). The hepatitis B virus. *Nature*; **317:** 489-495.

Tokino, T. and Matsubara, K. (1991). Chromosomal sites for hepatitis B virus integration in human hepatocellular carcinoma. *J. Virol.*; **65:** 6761-6764.

Tran, A., Kremsdorf, D., Capel, F., Housset, C., Dauguet, C., Petit, M.-A. and Brechot, C. (1991). Emergence of and takeover by hepatitis B virus (HBV) rearrangements in the pre-S/S and pre-C/C genes during chronic HBV infection. *J. Virol.*; **65:** 3566-3574.

Treichel, U., Meyer zum Büschenfelde, K.H., Dienes, H.P. and Gerken, G. (1997). Receptor mediated entry of hepatitis B virus particles into liver cells. *Arch. Virol.*; **142:** 493-498.

Tsai, S.-L., Chen, M.-H., Yeh, C.-T., Chu, C.-M., Lin, A.-N., Chiou, F.-H., Chang, T.-H. and Liaw, Y.-F. (1996). Purification and characterization of a naturally processed hepatitis B virus peptide recognized by CD8+ cytotoxic T lymphocytes. *J. Clin. Invest.*; **97:** 577-584.

Tswana, S., Chetsanga, C., Nystrom, L., Moyo, S., Nzara, M. and Chieza, L. (1996). A sero-epidemiological cross-sectional study of hepatitis B virus in Zimbabwe. *S. Afr. Med. J.*; **86:** 72-75.

Tur-Kaspa, R., Shaul, Y., Moore, D.D., Burk, R.D., Okret, S., Poellinger, L. and Shafritz, D. A (1988). The glucocorticoid receptor recognizes a specific nucleotide sequence in hepatitis B virus DNA causing increased activity of the HBV enhancer. *Virology*; **167:** 630-633.

Tuttleman, J.S., Pugh, J.C. and Summers, J.W. (1986). *In Vitro* experimental infection of primary duck hepatocyte cultures with duck hepatitis B virus. *J.Virol.;* **58:** 17-25.

Twu, J.S., Lee, C.H., Lin, P.M. and Schloemer, R.H. (1988). Hepatitis B virus suppresses expression of human beta-interferon. *Proc Natl Acad Sci USA*; **85:** 252-256.

Twu, J.S., Chu, K. and Robinson, W.S. (1989). Hepatitis B virus X gene activates κB-like enhancer sequences in the long terminal repeat of immunodeficiency virus. *Proc. Natl. Acad. Sci. USA;* **86:** 5168-5172.

Twu, J.S., Lai, M.-Y., Chen, D.-S. and Robinson, W.S. (1993). Activation of protooncogene C-*jun* by the X protein of hepatitis B virus. *Virology*; **192:** 346-350.

Tyrrell, B. (1996). *Her African Quest.* 1st ed. Western Cape: Lindlife Publishers CC.

Uchida, T., Aye, T.T., Shikata, T., Yano, M., Yatsuhashi, H., Koga, M. and Mima, S. (1994). Evolution of the hepatitis B virus gene during chronic infection in seven patients. *J. Med. Virol.;* **43:** 148-154.

Urban, S., Kruse, C. and Multhaup, G. (1999). A soluble form of the avian hepatitis B virus receptor. Biochemical characterization and functional analysis of the receptor ligand complex. *J. Biol. Chem.;* **274:** 5707-5715.

Usuda, S., Okamoto, H., Iwanari, H., Baba, K., Tsuda, F., Miyakawa, Y. and Mayumi, M., (1999). Serological detection of hepatitis B virus genotypes by ELISA with monoclonal antibodies to type-specific epitopes in the preS2-region product. *J. Virol. Meth*; **80:** 97-112.

Usuda, S., Okamoto, H., Tanaka, T., Kidd-Ljunggren, K., Holland, P.V., Miyakawa, Y. and Mayumi, M. (2000). Differentiation of hepatitis B virus genotypes D and E by ELISA using monoclonal antibodies to epitopes on the preS2-region product. *J. Virol. Meth*; **87:** 81-89.

Uy, A., Wunderlich, G., Olsen, D.B., Heermann, K.- H., Gerlich, W.H. and Thomssen, R. (1992). Genomic variability in the preS1 region and determination of routes of transmission of hepatitis B virus. *J. Gen. Virol.*; **73:** 3005-3009.

Valliammai, T., Echevarria, J.M., Leon, P., Tusets, C. and Harrison, J. (1995). Amplification and sequence analysis of the precore and core region of the HBV genome from sera of Spanish patients with HBV_2–like infections. *J. Med. Virol.;* **46:** 375-379.

van Hattum, J., Schreuder, G.M. and Schalm, S.W. (1987). HLA antigens in patients with various courses after hepatitis B virus infection. *Hepatology*; **7:** 11-14.

Vardas, E., Mathai, M., Blaauw, D., McAnerney, J., Coppin, A. and Sim, J. (1999). Preimmunization epidemiology of hepatitis B virus infection in South African children. *J. Med. Virol.*; **58:** 111-115.

Vardas, E. (1999): *The epidemiology and molecular characteristics of hepatitis B virus infection in children from a hyperendemic area of South Africa, including a field trial of the hepatitis B immunisation programme and its impact on infection in this population.* A dissertation submitted in partial fulfilment of the requirements for the degree Master of Medicine (Virology), Faculty of Medicine, University of the Witwatersrand, Johannesburg.

Vaudin, M., Wolstenholme, A.J., Tsiquaye, K.N., Zuckerman, J. aand Harrison, T.J. (1988). The complete nucleotide sequence of the genome of a hepatitis B virus isolated from a naturally infected chimpanzee. *J. Gen. Virol.;* **69:** 1383-1389.

Walker, M.E., Szmuness, W., Stevens, C.E. and Rubinstein, P. (1981). Genetics of anti-HBs responsiveness. I. HLA-DR7 and non-responsiveness to hepatitis vaccination (Abstract). *Transfusion;* **21:** 601.

Wang, G.-H. and Seeger, C. (1993). Novel mechanism for reverse transcription in hepatitis B viruses. *J. Virol.*; **67:** 6507-6512.

Wang, H.P. and Rogler, C.E. (1991). Topoisomerase I-mediated integration of hepadnavirus DNA *in vitro*. *J. Virol.*; **65:** 2381-2392.

Wang, X.W., Forrester, K., Yeh, H., Feitelson, M.A., Gu, J.-R. and Harrris, C.C. (1994). Hepatitis B virus X protein inhibits p53 sequence-specific DNA binding, transcriptional activity and association with transcription factor ERCC3. *Proc. Natl. Acad. Sci. USA;* **91:** 2230-2234.

Warren, K.S., Heeney, J.L., Swan, R.A., Heriyanto, Verschoor, E.J. (1999). A new group of hepadnaviruses naturally infecting orangutans (*Pongo pygmaeus*). *J. Virol.*; **73:** 7860-7865.

Wasylyk, B. and Chambon, P. (1981). A T to A base substitution and small deletions in the conalbumin TATA box drastically decrease specific *in vitro* transcription. *Nucleic Acids Res.*; **9:** 1813-1824.

Waters, J., Kennedy, M., Voet, P., Hauser, P., Petre, J., Carman, W. and Thomas, H.C. (1992). Loss of a common '*a*' determinant of hepatitis B surface antigen by a vaccine escape mutant. *J. Clin. Invest.*; **90:** 2543-2547.

Wei, Y., Etiemble, J., Fourel, G., Vitvitski-Trepo, L. and Buendo, M.-A. (1995). Hepadna virus integration generates virus-cell cotranscripts carrying 3' truncated X genes in human and woodchuck liver tumours. *J. Med. Virol.;* **45:** 82-90.

Whitten, T.M., Quets, A.T. and Schloemer, R.H. (1991). Identification of the hepatitis B virus factor that inhibits expression of the beta interferon gene. *J. Virol.*; **65:** 4699-4704.

Whittle, H.C., Inskip, H., Hall, A.J., Mendy, M., Downes, R. and Hoares, S. (1991). Vaccination against hepatitis B and protection against chronic viral carriage in The Gambia.

Lancet; **337:** 747-750.

Will, H., Reiser, W, Weimer, T., Pfaff, E, Buscher, M., Sprengel, R., Cattaneo, R. and Schaller, H. (1987). Replication strategy of human hepatitis B virus. *J. Virol.*; **61:** 904-911.

World Health Organization (October, 1996). Newsletter of the WHO.

Xu, Z., Jensen, G. and Yen, T.S.B. (1997). Activation of hepatitis B virus S promoter by the viral large surface protein via induction of stress in the endoplasmic reticulum. *J. Virol.*; **71:** 7387-7392.

Yaginuma, K. and Koike, K. (1989). Identification of a promoter region for 3.6-kilobase mRNA of hepatitis B virus and specific cellular binding protein. *J. Gen.Virol.*; **63:** 2914-2920.

Yang, Z., Lauder, I.J. and Lin, H.J. (1995a). Molecular evolution of the hepatitis B virus genome. *J. Mol. Evol.;* **41:** 587-596.

Yang, Z. and Summers, J. (1995b). Illegitimate replication of linear hepadnavirus DNA through nonhomologous recombination. *J. Virol.;* **61:** 904-911.

Yeh, C.-T., Wong, S.W., Fung, Y.-K. and Ou, J.-H. (1993). Cell cycle regulation of nuclear localization of hepatitis B virus core protein. *Proc. Natl. Acad. Sci. USA*; **90:** 6459-6463.

Yen, T.S.B. (1996). Hepadnaviral X protein: review of recent progress. *J. Biomed. Sci.;* **3:** 10-30.

Yoffe, B., Burns, D.K., Bhatt, H.S. and Combes, BI (1990). Extrahepatic hepatitis B virus DNA sequences in patients with acute hepatitis B infection. *Hepatology;* 12: 187-192.

Yuan, T.T.-T., Lin, M.-H., Qiu, S.M. and Shih, C. (1998). Functional characterization of

naturally occurring variants of human hepatitis B virus containing the core internal deletion mutation. *J. Virol.*; **72:** 2168-2176.

Yuan, T.T. and Shih, C. (2000). A frequent naturally occurring mutation (P130T) of human hepatitis B virus core antigen is compensatory for immature secretion phenotype of another frequent variant. *J. Virol.*; **74:** 4929-4932.

Yuasa, R., Takahashi, K., Dien, B.V., Binh, N.H., Morishita, T., Sato, K., Yamamoto, N., Isomura, S., Yoshioka, K., Ishikawa, T., Mishiro, S. and Kakumu, S. (2000). Properties of hepatitis B virus genome recovered from Vietnamese patients with fulminant hepatitis in comparison with those of acute hepatitis. *J. Med. Virol.*; **61:** 23-28.

Yuh, C.-H., Chang, Y.-L. and Ting, L.-P. (1992). Transcriptional regulation of precore and pregenomic RNAs of hepatitis B virus. *J. Virol.*; **66:** 4073-4084.

Zhang, X., Zoulim, F., Habersetzer, F., Ziong, S. and Trepo, C. (1996). Analysis of hepatitis B virus genotypes and pre-core region variability during interferon treatment of HBe antigen negative chronic hepatitis B. *J. Med. Virol.*; **48:** 8-16.

Zollner, B., Petersen, J., Schroter, M., Laufs, R., Schoder, V. and Feucht, H.H. (2001). 20-fold increase in risk of lamivudine resistance in hepatitis B virus subtype *adw. Lancet;* **357:** 934-935,