

https://dspace.org/

Characterization of mutants and splice variants of hepatitis B virus isolated from South African black hepatocellular carcinoma patients

Skelton, Michelle

2010-02-15T09:36:37Z
http://hdl.handle.net/10539/7511

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

CHARACTERIZATION OF MUTANTS AND SPLICE VARIANTS OF HEPATITIS B VIRUS ISOLATED FROM SOUTH AFRICAN BLACK HEPATOCELLULAR CARCINOMA PATIENTS.

Michelle Skelton

Thesis submitted in compliance for the requirements for the degree of Doctor of Philosophy in the Faculty of Health Sciences at the University of the Witwatersrand

Johannesburg 2009

DECLARATION

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

…….

……..day of…….(month), 2009

DEDICATION

I dedicate this thesis to my dear and late grandparents

Lydia Deist, Mary Skelton and Frederick Deist.

PRESENTATIONS

1. Conference: Cancer Research UK Beatson International Cancer Conference, pg 115, Scotland, UK.

   Poster: Michelle Skelton, Gerald C Kimbi, Anna Kramvis and Michael C. Kew (2005). Characterization of hepatitis B virus mutants in hepatocellular carcinoma in southern Africa.
2. Workshop: South African-Swedish Research Partnership (SIDA/NRF) Workshop Phylogenetic analysis of Hepatitis viruses found in Southern Africa relative to those from other geographic regions of the world. UWC, Cape Town, South Africa (2006).

   Presentation: Characterization of hepatitis B virus mutants in hepatocellular carcinoma in southern Africa.
3. MRC review of the Molecular Hepatology Research Unit in Johannesburg (27 June 2006).

   Presentation: Characterization of hepatitis B virus mutants in hepatocellular carcinoma in southern Africa.
4. Symposium: Hepatitis B Virus Genotypes ~ from an academic question to the clinic (July 2008).

   Presentation: Subgenotype A1 splice variants and mutants isolated from serum of South Africa hepatocellular carcinoma patients.

ABSTRACT

Hepatitis B virus (HBV) infection is endemic in Africa. As many as 98% of black Africans are infected during their lives and about 10% (65 million) have chronic HBV infection, which is the cause of 70-80% of all hepatocellular carcinoma (HCC) cases. Despite this high prevalence of HBV and the high incidence of HCC in Africa, relatively few complete HBV genomes from African HCC cases have been deposited in international data bases. In order to gain a clearer understanding of the role of genetic variants and mutants in the development of HCC, the complete genomes of HBV isolated from southern African HCC patients were amplified and molecularly characterized. HBV DNA was extracted from the serum forty HBsAg-positive HCC patients. Twenty six complete genomes were successfully amplified, cloned and sequenced from nine HCC patients.

Phylogenetic analyses of the complete genomes and the individual open reading frames of HBV isolates from the HCC patients, led to the classification of all the isolates within subgenotype A1. No isolates belonging to subgenotype A2 and genotype D were identified, even though these genotypes/subgenotypes have been shown to circulate in South Africa. Three patients contained the uncommon combination of serological subtype *ayw*1 in the subgenotype A1 strain. This combination has been found previously in South Africa and the Phillipines.

Seventy-eight percent of the patients carried HBV strains with the double basic core promoter (BCP) mutation (1762T/1764A), previously shown to reduce HBeAg expression. Furthermore, complete genome sequence analysis has revealed a complex combination of mutations, which include at least three or five of these residues 1753C1762T1764A1766T1768A1809T1812T occurring as the dominant

HBV strains isolated from 5/9 HCC patients. These mutations have previously been shown to regulate gene expression at various levels, to enhance viral replication and simultaneously decrease HBeAg expression.

All five HBV genomes isolated from one patient contained novel complex BCP rearrangements, which introduced 2 HNF1 and 1 putative HNF3 transcription factor binding sites. These mutations can enhance viral replication and simultaneously abolish HBeAg expression at a transcriptional level. Furthermore, truncated core proteins would be expressed from 4/5 isolates and none would express wild-type HBx. Several mutations were identified in the pre-S/S genes of 2/5 isolates, which would result in the expression of novel 3' truncated medium surface proteins ($MHBs^t$) and large surface proteins ($LHBs^t$). The majority of the mutations would contribute to hepatocyte pathogenesis and transformation by activating cell proliferating pathways.

Two patients also contained rare HBV variants not previously identified in HBV strains from southern Africa. These included an HBV splice variant and a poly (dA) variant from patient 10 and patient 6, respectively. These variants occurred in combination with other isolates within the respective patients.

The envelope genes were characterised in a total of 18 HCC patients, the pre-S gene of HBV contained deletions in 72% of the patients. Deletions across pre-S1/pre-S2, pre-S2 initiation codon mutations with internal deletions, and S gene nonsense mutations were prevalent. Mutated envelope proteins have been shown to accumulate within the hepatocyte endoplasmic reticulum (ER) and are a characteristic histopathological hallmark of HCC known as ground glass

hepatocytes. HBV induced ER stress has been shown to dysregulate several cell cycle regulatory pathways, which contribute to HCC.

In addition several novel $LHBs^t$ and $MHBs^t$ have been described. These potential transactivators require further investigation. The HBV mutations described in this study have been associated with increased risk for HCC.

Despite the obvious heterogeneity HBV displays within and between patients, there are common characteristics shared between the HBV variants which emerge during the development of HCC. These include the BCP and pre-C (1753C1762T1764A1766T1768A1809T1812T) mutations and the pre-S/S mutations. These mutations are able to affect HBV replication and gene expression, and may work synergistically to promote liver dysfunction and HCC.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Professor Michael Kew. His enthusiasm towards my project has been a source of encouragement. He has taken much time and care in editing my thesis. He has provided invaluable criticism through his combined experience and expertise in clinical and molecular hepatitis research. It is an honour to be associated with such an accomplished leader in the field of hepatitis and cancer research.

I would also like to thank my co-supervisor, Professor Anna Kramvis, affectionately known as, Anna. This project was her brainchild. Anna is a visionary in the field of molecular evolution of HBV and I am proud to be her student. Thank you for your idealism and pursuit for excellence. You have encouraged me to thrash out every aspect of this thesis and I appreciate that.

I would like to thank the staff and friends in the Molecular Hepatology Research Unit at the University of the Witwatersrand. A special thanks to Roshni Desai for always being pleasant and carrying out her responsibilities timeously so that projects could run smoothly. Alice Chen, my dear friend thank you for your support. To Gerald Kimbi whose work laid the foundation for my research.

A warm thank you to my friends and colleagues at the University of Cape Town where I spent a considerable time writing and analysing data. Thank you for embracing me and helping me through a difficult time. A special thanks to Professor Iqbal Parker, Dr Sharon Prince, Widaad Zemanay, Catherine Whibley, Dr Virna Leaner and Gama Bandawe.

A huge thanks to my Dad for his love, patience, financial support and just being there. Mom, you appear to be the number one fan of this huge endeavour called my PhD. Thank you for all the faith you have in me, it has truly made all the difference. Thank you both for caring so much for me.

Thank you to all my family and friends who have always believed in me, it is much appreciated.

I would like to thank the National Research Foundation and the Poliomyelitis Research Foundation for their financial support.

TABLE OF CONTENTS

CHAPTER 1

1.0	**INTRODUCTION**	1
1.1	THE FAMILY OF *HEPADNAVIRIDAE*	4
1.2	VIRAL STRUCTURE	4
1.3	VIRAL GENOME	6
1.4	HBV TRANSCRIPTS	8
1.5	VIRAL GENE PRODUCTS	9
1.5.1	SURFACE PROTEINS	9
1.5.2	CORE GENE PRODUCTS	15
1.5.3	POLYMERASE	16
1.5.4	HBx PROTEIN	17
1.6	VIRAL *CIS*-ELEMENT	17
1.6.1	PROMOTERS AND ENHANCERS	17
1.6.2	THE ENCAPSIDATION SIGNAL	22
1.6.3	THE POLYADENYLATION SIGNAL	22
1.7	HBV LIFE CYCLE	23
1.8	IMMUNO-PATHOGENESIS OF HBV INFECTION	28
1.8.1	TRANSMISSION	28
1.8.2	PATHOGENESIS OF HBV INFECTION	28
1.8.3	ACUTE HEPATITIS	29
1.8.4	CHRONIC (PERSISTENT) HEPATITIS	30
1.8.5	HEPATOCELLULAR CARCINOMA	32
1.9	HBV GENOTYPES: MOLECULAR CHARACTERISTICS AND GLOBAL DISTRIBUTION	33
1.9.1	GENOTYPE A	37
1.9.2	GENOTYPE B	37
1.9.3	GENOTYPE C	38
1.9.4	GENOTYPE D	38
1.9.5	GENOTYPE E	38
1.9.6	GENOTYPE F	39
1.9.7	GENOTYPE G	39
1.9.8	GENOTYPE H	40
1.10	HBV SEQUENCE VARIATION	40
1.10.1	HBV GENOTYPES AND DISEASE PROGRESSION	41
1.10.2	HBV MUTANTS AND DISEASE PROGRESSION	43
1.10.2.1	CORE PROMOTER AND ENHANCER II REGION	43

1.10.2.2 PRECORE AND CORE GENE 45
1.10.2.3 X GENE 46
1.10.2.4 POLYMERASE REGION 48
1.10.2.5 THE PRE-S AND SURFACE REGION 49
1.11 PRE-S FAMILY OF TRANSACTIVATORS 53
1.12. SPLICE VARIANTS AND VARIANTS WITH INTERNAL POLY (dA) TAILS 54
1.13 HBV DNA INTEGRATION 55
1.14 THE HBx TRANSACTIVATOR 55
1.15 RECOMBINANTS 56
1.16 HBV SEQUENCE VARIATION IN SOUTHERN AFRICA 57
1.17 AIMS AND ORGANIZATION OF STUDY 60

CHAPTER 2

2.0. MATERIALS AND METHODS 63
2.1 SUBJECTS AND SERUM SAMPLES 63
2.2 DNA EXTRACTION, AMPLIFICATION, CLONING AND SEQUENCING 65
2.2.1 DNA EXTRACTION 65
2.2.2 DNA AMPLIFICATION 65
2.2.2.1 COMPLETE GENOME AMPLIFICATION 65
2.2.2.2 SUB-GENOMIC AMPLIFICATION 66
2.2.2.3 SEMI-NESTED POLYMERASE CHAIN REACTION (PCR) 66
2.2.3 DETECTION OF AMPLICONS 69
2.2.4 CLONING OF AMPLICONS 69
2.3 DNA SEQUENCING 72
2.3.1 MANUAL SEQUENCING 72
2.3.1.1 PREPARATION OF PCR TEMPLATES 72
2.3.1.2 PREPARATION OF PLASMID DNA 72
2.3.1.3 PRIMER ANNEALING AND LABELLING REACTION 72
2.3.2 AUTOMATED SEQUENCING 73
2.4 PHYLOGENETIC ANALYSES 75
2.5 CONSTRUCTION OF SURFACE GENE EXPRESSION VECTORS 75
2.6 PREPARATION OF THE PRE-S/S GENE FRAGMENTS 76
2.7 LIGATION REACTION 77
2.8 SUB-CLONING STRATEGY 77

2.9 CELL CULTURE 79
2.9.1 CELL LINE 79
2.9.2 SUBCULTURE AND STOCKS 79
2.9.3 MYCOPLASMA TEST 80
2.10 CALCIUM PHOSPHATE TRANSFECTION 81
2.11 DETECTION OF B-GALACTOSIDASE 82
2.12 DOXORUBICIN TREATMENT 82
2.13 PROTEIN PURIFICATION 83
2.13.1 CELL LYSATE PREPARATION 83
2.13.2 PROTEIN QUANTITATION 83
2.14 WESTERN BLOT ANALYSIS 84
2.14.1 SODIUM DODECYL SULFATE POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE) 84
2.14.2 TRANSFER 84
2.14.3 PONCEAU S (MEMBRANE) AND COOMASSIE BLUE (GEL) STAINS 84
2.14.4 PRIMARY ANTIBODY 85
2.14.5 SECONDARY ANTIBODY 85
2.14.6 MEMBRANE STRIPPING 86
2.15 CASPASE 3 ASSAY 86
2.16 PLASMID DNA EXTRACTIONS FROM BACTERIAL CELLS 87
2.17 PREPARATION OF COMPETENT BACTERIAL CELLS 88

CHAPTER 3

3.0 RESULTS 89
3.1 SEQUENCE CHARACTERISTICS AND PHYLOGENETIC ANALYSIS OF HBV FROM BLACK SOUTHERN AFRICAN HEPATOCELLULAR CARCINOMA PATIENTS 89
3.1.1 AMPLIFICATION, CLONING AND SEQUENCING 89
3.1.2 PHYLOGENETIC ANALYSIS 91
3.1.3 COMPARISON OF AMINO ACID SEQUENCES OF SUBGENOTYPE A1 HBV ISOLATES FROM HCC PATIENTS WITH THOSE FROM ASYMPTOMATIC, CHRONIC CARRIERS, ACUTE HEPATITIS PATIENTS AND FULMINANT HEPATITIS PATIENTS 100
3.1.4 COMPARISON OF NUCLEOTIDE SEQUENCES OF SUBGENOTYPE A1 HBV ISOLATED FROM HCC PATIENTS WITH ISOLATES FROM ASYMPTOMATIC CHRONIC CARRIERS, ACUTE AND FULMINANT HEPATITIS PATIENTS BELONGING TO SUBGENOTYPE A1 105

3.2 SIMULTANEOUS ALTERATIONS IN THE BASIC CORE PROMOTER, CORE GENE AND SURFACE GENE 108
3.2.1 PCR AND CLONING 108
3.2.2 FULL GENOME STRUCTURE 111
3.2.3 STRUCTURE OF THE BCP MUTATION 111
3.2.3.1 DELETION, INSERTION AND DUPLICATION 111
3.2.3.2 INITIATION SITES AND TA-RICH MOTIFS 112
3.2.3.3 TRANSCRIPTION FACTOR BINDING SITES 112
3.2.4 X-GENE MUTATIONS AND PREDICTED PROTEIN EXPRESSION 113
3.2.5 CORE GENE MUTATIONS AND PREDICTED PROTEIN EXPRESSION 113
3.2.6 SURFACE GENE MUTATIONS 114
3.3 IDENTIFICATION AND MOLECULAR CHARACTERIZATION OF RARE HBV VARIANTS IN HEPATOCELLULAR CARCINOMA 126
3.3.1 HBV ISOLATED FROM PATIENT 6 128
3.3.1.1 GENOME HETEROGENEITY 131
3.3.1.2 STRUCTURE OF TYPE I GENOME 131
3.3.1.3 PREDICTED PROTEIN SEQUENCE 132
3.3.2 HBV ISOLATED FROM PATIENT 10 133
3.3.2.1 GENOME HETEROGENEITY 136
3.3.2.2 STRUCTURE OF THE TYPE I GENOME 136
3.3.2.3 PREDICTED PROTEIN SEQUENCE 137
3.3.3 PHYLOGENETIC ANALYSIS 138
3.4 CHARACTERISATION OF PRE-S MUTATIONS IN HBV ISOLATED FROM HCC PATIENTS 140
3.4.1 PRE-S MUTATIONS 140
3.4.2 PRE-S MUTATION PATTERN AND HETEROGENEITY 141
3.4.3 C-TERMINALLY (3') TRUNCATED PRE-S/S GENE MUTANTS 143
3.4.4 PHYLOGENETIC ANALYSIS 142
3.5 FUNCTIONAL ANALYSIS OF MUTANT PRE-S/S GENES 152
CHAPTER 4
4.0 DISCUSSION 157
4.1 SEQUENCE CHARACTERISTICS AND PHYLOGENETIC ANALYSIS OF HBV FROM BLACK SOUTHERN AFRICAN HEPATOCELLULAR CARCINOMA PATIENTS 157

4.1.1 HBV HETEROGENEITY....157
4.1.2 GENETIC RELATEDNESS AND PREDICTIVE MARKERS FOR HCC....158
4.1.3 SUBGENOTYPE A1 CHARACTERISTICS AND SIGNATURE SEQUENCES....165
4.1.4 MUTATIONS IN HBV PRE-C GENE AND BCP IN HCC IN SOUTHERN AFRICA....168
4.1.5 CONCLUSION....173
4.2 SIMULTANEOUS ALTERATIONS IN THE BASIC CORE PROMOTER, CORE GENE AND SURFACE GENE....174
4.2.1 HBV HETEROGENEITY WITHIN PATIENT 18....174
4.2.2 PREVALENCE OF THE COMPLEX BCP MUTATIONS....175
4.2.3 HOMOLOGOUS RECOMBINATION MEDIATED BY TOPOISOMERASE I....178
4.2.4 BCP REARRANGEMENTS....184
4.2.5 THE CORE GENE AND PRODUCTS....187
4.2.6 THE PRE-S/S GENE AND PRODUCTS....188
4.2.7 THE X-GENE AND PRODUCTS....189
4.2.8 CUMULATIVE EFFECT OF MULTIPLE MUTATIONS IN ALL THE ORFS....190
4.2.9 CONCLUSION....191
4.3 IDENTIFICATION OF RARE HBV VARIANTS IN HCC....192
4.3.1. HBV HETEROGENEITY IN PATIENT 6 AND PATIENT 10....192
4.3.2. THE POLY (dA) VARIANT....193
4.3.2.1 GENOME STRUCTURE AND REPLICATION STRATEGY....193
4.3.2.2 RECOMBINANT PROTEINS....194
4.3.3 HBV SPLICE VARIANT....196
4.3.3.1 GENOME STRUCTURE....196
4.3.3.2 RECOMBINANT PROTEINS....197
4.3.3.3 HBV SPLICE VARIANTS AND HCC....199
4.3.4 CONCLUSION....199
4.4 CHARACTERISATION OF PRE-S/S GENE MUTANTS IN HCC IN BLACK SOUTHERN AFRICANS....201
4.4.1 PREVALENCE AND DISEASE....201
4.4.2 THE PRE-S1 AND PRE-S2 REGIONS....202
4.4.3 PRE-S1/S2 MUTANTS AND GGHS....205

4.4.4 NOVEL TRANSACTIVATORS.....205
4.4.5 CELL CYCLE REGULATORY PATHWAYS AFFECTED BY ΔPRE-SLHBS/MHBS.....208
4.4.6 CONCLUSION 209

4.5 MUTANT PRE-S/S GENES AND THE CELL CYCLE REGULATORY PATHWAYS 210

4.5.1 CONCLUSION 215

CHAPTER 5

5.0 GENERAL CONCLUSION.....216

APPENDICES 220
APPENDIX A: SOLUTIONS AND REAGENTS.....220
APPENDIX B: Figure B1 225
REFERENCES 233

LIST OF FIGURES

Figure 1.1 Geographical distribution of chronic hepatitis B virus Infection 3

Figure 1.2A Diagrammatic representation of a Dane particle or complete HBV virion. The envelope is made up of LHBs, MHBs and SHBs 5

Figure 1.2B Filamentous sub-viral particle, which are mainly made up of SHBs and MHBs. 5

Figure 1.2C Spherical viral particle is composed mainly of SHBs and MHBs. 5

Figure 1.3 Organization of the HBV genome…….. 7

Figure 1.4 A schematic representation of the HBV surface proteins. 12

Figure 1.5 Transmembrane topology of the HBV envelope proteins. 14

Figure 1.6 Core promoter/Enhancer II region 21

Figure 1.7A The HBV life cycle. 25

Figure 1.7B Representation of viral replication 26

Figure 1.8 Outline of HBV-induced disease and the prevalences of the various sequelae characteristic of southern Africans 29

Figure 1.10 The ‘a’ determinant of the HBsAg occurring within the major hydrophilic loop 52

Figure 2.1 Flow diagram of methodology 64

Figure 3.1 Dendrogram based on the complete genomes of 24 genotype A1 isolates from HCC and representative sequences from subgenotype A1, A2 and the 7 remaining genotypes 94

Figure 3.2 Dendrogram based on the complete polymerase genes of 24 subgenotype A1 isolates from HCC and representative sequences from subgenotype A1, A2 and the 7 remaining genotypes 95

Figure 3.3 Dendrogram based on the complete pre-C/Core genes of 24 subgenotype A1 isolates from HCC and representative sequences from subgenotype A1, A2 and the 7 remaining genotypes 96

Figure 3.4 Dendrogram based on the complete surface genes of 24 subgenotype A1 isolates from HCC and representative sequences from subgenotype A1, A2 and the 7 remaining genotypes 97

Figure 3.5 Dendrogram based on the complete X-genes of 24

subgenotype genotype A1 isolates from HCC and representative sequences from subgenotype A1, A2 and the 7 remaining genotypes 98
Figure 3.6 Comparison of amino acid residues of S, polymerase and X-ORFs of subgenotype A1 isolates from HCC patients, asymptomatic chronic carriers, acute hepatitis patients and fulminant hepatitis patients....101
Figure 3.7 Comparison of nucleic acid sequences of the *cis*-acting elements of subgenotype A1 isolates from HCCs, asymptomatic chronic carriers and acute hepatitis patients. Dots indicate amino acid identity....103
Figure 3.8 PCR amplicons and plasmid DNA restriction digest of HBV clones derived from patient 18 109
Figure 3.9 Full-genome structure of five HCC 18 clones with genome lengths determined by complete sequence analysis 110
Figure 3.10 Nucleotide sequence of the BCP (nt 1741-nt 1849) genotype A 115
Figure 3.10.A Initiation sites of the precore (pre-C) mRNA and pregenomic (pg) RNA are shown as horizontal arrows and initiators (Inr) are shown in italics. The BCP contains four AT-rich domains (TA1-TA4).The positions of the 13 nucleotide deletion and the 45 nucleotide insertion are indicated. 115
Figure 3.10.B Reorganization of the AT-rich domains in BCP for all five clones. TA2 motifs are shown in bold and a new motif TATTA and TTAAATATTA are underlined 115
Figure 3.10.C Two 27 nucleotide repeats are underlined and are made up of a recombination of overlapping adjacent sequences, demonstrated by the overlapping triangles 115
Figure 3.11 Nucleotide sequence of the Core promoter showing transcription factor binding sites....117
Figure 3.12 Predicted X protein amino acid sequences expressed from mutants....119
Figure 3.13 Structure of the core gene mutation....120
Figure 3.14 Chromatogram produced following automated sequencing of the core gene of HBV isolated from patient 18....122
Figure 3.15 HBV full-length PCR amplicons from patient DNA resolved on 1% agarose gels stained with ethidium bromide. 127
Figure 3.16 *Pvu*I restriction digestion products of cloned HBV resolved on a 1%

agarose gel. 128
Figure 3.17 Full-genome structure of 6 clones derived from HCC patient 18 determined using sequence analysis. 129
Figure 3.17.A At the top is a schematic presentation of the genome structure with four overlapping reading frames and the relevant *cis*-elements, including the basic core promoter (BCP), direct repeat 1 (DR1), direct repeat 2 (DR 2), polyadenylation signal (poly A). The poly (dA) tail is shown in the rectangular box, followed by a large deletion illustrated with the broken line and dots 129
Figure 3.17.B Chromatograph of clone 6.1 HBV DNA sequence obtained using primer T3. 129
Figure 3.18 Amino acid sequence of the recombinant protein generated for the core-polymerase ORF of the poly (dA) variant. 132
Figure 3.19 Molecular structure of three HBV genomes isolated from HCC patient 10 determined by complete sequence analysis. 134
Figure 3.19.A At the top is a schematic presentation of the genome structure with four overlapping reading frames and the relevant *cis*-elements. The basic core promoter (BCP), direct repeat 1 (DR1) and direct repeat 2 (DR 2). The large deletion is illustrated with the broken line and dots. 134
Figure 3.19B Deletion boundary sequences of genome 10.17 were aligned to the consensus sequences for the 5' – splice site (MAG/GTRAGT), the 3' splice site (YAG/G), and the branch site (YNYYRAY) (Jackson, 1991).The branch site and polypyrimidine tract are indicated, non-negotiable nucleotides are shown in bold 134
Figure 3.20 Amino acid sequence of the recombinant protein generated for the core-polymerase ORF of the splice variant. 137
Figure 3.21 Dendrogram based on the complete X-genes of 26 subgenotype genotype A1 isolates from HCC and representative sequences from subgenotype A1, A2 and the 7 remaining genotypes 139
Figure 3.22.A Amino acid sequence alignment of the preS1 region 144
Figure 3.22.B Amino acid sequence alignment of the preS1 region. 145
Figure 3.23 Schematic presentation of pre-S mutation patterns found in the southern African HBV subgenotype A1 isolates. 149

Figure 3.24 Schematic presentation of LHB, LHB 3' truncated mutants (LHBs^{t}) and MHB 3' truncated mutants (MHBs^{t}). 150

Figure 3.25 Dendrogram based on the complete surface genes of 47 subgenotype genotype A1 isolates from HCC and representative sequences from subgenotype A1, A2 and the 7 remaining genotypes. 151

Figure 3.26 Schematic presentation of the preS/S gene constructs (pre-S variants). 154

Figure 3.27A Transfection efficiency measure in parallel using X-gal staining 155

Figure 3.27B $p21^{Waf-1}$ western blot analysis. 155

Figure 3.27C Caspase 3 activity 155

Figure 4.1. Proposed model of Topoisomerase I-mediated recombination in the HBV strains isolated from patient 18 180

.

LIST OF TABLES

Table 1.1. HBV Transcripts....8

Table 1.2. Genotype Variations....34

Table 1.3. Geographical Distribution of Genotypes and Subgenotypes....36

Table 2.1. PCR Primers and conditions used for amplification....67

Table 2.2. Primers used for sequencing....74

Table 2.3. Ligation Reaction Components....77

Table 2.4. Surface gene expression vectors....79

Table 2.5. Calcium Phosphate transfection components....81

Table 2.6. Antibodies used in the Western Blot Analysis....86

Table 3.1.1. PCR and Cloning HBV Data of HCC Patients....90

Table 3.1.2. Mean pair-wise DNA divergence (%) of complete genome sequences and individual ORFs of subgenotype A1 in different patient groups....93

Table 3.1.3. Mutations within the *cis*-acting regulatory elements found in subgenotype A1 HCC patients....106

Table 3.1.4. Core promoter and precore mutations occurring in complete genomes of HCC patients in the current study....107

Table 3.2.1. Mutations in the cloned HBV genomes detected by sequence analysis....124

Table 3.4.1. Deletions in the pre-S domain and nucleotide position....147

Table 3.4.2. Mutations within the Virion Morphogenesis Domain (aa 103-124) 148

LIST OF ABBREVIATIONS

ε	encapsidation signal
aa	amino acid
ALT	Alanine aminotransferase
APC	Antigen presenting cell
ASC	Asymptomatic chronic carrier
AS	Asymptomatic chronic carrier (pertaining to figures)
ASHV	Arctic squirrel hepatitis B virus
bp	base pair
BCP	Basic core promoter
BQW	Best quality water
cccDNA	Covalently closed circular DNA
COR	Cohesive overlap region
COUP-TF	Chicken ovalbumin upstream promoter transcription factor
CURS	Core upstream regulatory sequence
DHBV	Duck hepatitis B virus
DNA	Deoxyribonucleic Acid
DNAML	DNA maximum likelihood
DR1	Direct repeat one
DR2	Direct repeat two
ERK	Extracellular signal-regulated kinase
ENHI	Enhancer one
ENHII	Enhancer two
ER	Endoplasmic reticulum
ESLD	End stage live disease

GGH	Ground glass hepatocyte
GHSV	Ground squirrel hepatitis B virus
HBcAg	Hepatitis B core antigen
HBeAg	Hepatitis B e antigen
HBsAg	Hepatitis B surface antigen
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HHBV	Heron hepatitis B virus
HNF	Hepatocyte nuclear factor
IFN	Interferon
IPTG	isopropyl-beta-D-thiogalactopyranoside
kb	kilobase
LB	Luria-Bertini broth
LHBs	Large surface proteins
$LHBs^t$	Truncated large surface proteins
MAPK	Mitogen activated protein kinase
MHBs	Medium surface proteins
$MHBs^t$	Truncated medium surface proteins
mRNA	messenger RNA
Myr	Myristoylation
NEIGHBOR	neighbour-joining
NG	N-glycosylation
NRE	Negative regulatory element
OG	O-glycosylation
ORF	Open reading frame

PCR	Polymerase chain reaction
PK	Protein kinase
PKC	Protein kinase C
Pol	polymerase
Poly A	polyadenylation signal
PPAR α	peroxisome proliferator activated receptor α
PHYLIP	phylogeny inference package
RT	reverse transcriptase
RXR α	retinoid X receptor α
SHBs	Small (major) surface proteins
TBP	TATA binding protein
TP	terminal protein
Top I	Topoisomerase I
TR2	Human testicular receptor 2
URR	Upstream regulatory element
UV	Ultraviolet
WHO	World Health Organization
WHV	Woodchuck hepatitis B virus
WMHBV	Woolly Monkey hepatitis B virus
YMDD	Tyrosine, Methionine, Aspartic acid, Aspartic acid
YIDD	Tyrosine, Isoleucine, Aspartic acid, Aspartic acid
YVDD	Tyrosine, Valine, Aspartic acid, Aspartic acid

CHAPTER 1

1.0 INTRODUCTION

According to the most recent World Health Organization estimates, at least 2 billion people worldwide have been infected with the hepatitis B virus (HBV), and approximately 350 million are chronic carriers of the virus (WHO, 2002). Estimating HBV prevalences throughout the world has revealed regional differences (figure 1.1): prevalences vary from >8% (high) to 2-8% (intermediate) to <2% (low). The clinical manifestations of HBV infection range from an asymptomatic carrier state to acute self-limited or fulminant hepatitis to chronic hepatitis, that may progress to cirrhosis, or hepatocellular carcinoma (HCC) or both (Alberti et al., 2005). Globally 1.5 million deaths are attributed to HBV-related diseases each year, about 255 000 of these occurring in Africa (Kew, 2006), where as many as 98% of sub-Saharan black Africans are infected with the virus at some time during their lives and the virus is hyperendemic with an estimate of 65 million chronic HBV carriers (Kew, 2006; Kramvis and Kew, 2007a). Sub-Saharan Africa has an average carrier rate of 10.4% (range 5 - 20%). The HBV rates vary from 19.7% in Zimbabwe in southern Africa to 6.5% in Tunisia in North Africa (Kew, 1992; Kew, 2006). Small pockets of higher incidences (>35%) have also been documented (Kew, 2006). Lifetime exposure rates to the virus range from 57% in Nigeria to 98% in Namibia, with an average figure of 75% (Kew, 2006). The two sexes are equally likely to be infected, although chronic viral carriage is about twice (1.4-2.5:1) as common in males (Kew, 1992; Kew, 2006; Kramvis and Kew, 2007a). Similar high rates of chronic HBV infection are present in the Asian Pacific region (WHO, 2000).

Rural dwellers have a higher chronic HBV carrier rate compared to urban dwellers in most parts of Africa. For example, rural Kenyans have a carrier rate of 12% compared to 5.3% in Nairobi the Kenyan capital city (Bowry, 1983). Not only is this of epidemiological significance, but combined with poor access to health care services in rural areas, it means that rural communities generally shoulder a heavier disease burden than their urbanized counterparts.

The high HBV carrier rates in endemic regions and the resulting high risk of development of HCC is the result of a high rate of infection in infancy and early childhood. Two transmission patterns have been described based on geographical and virological differences. Horizontal transmission during early childhood predominates in Africa (Kramvis and Kew, 2007a), whereas perinatal transmission predominates in the Asian-pacific region. Perinatal transmission occurs in children born to highly infectious HBV e antigen-positive mothers. The percentage of highly infectious carrier mothers is far higher in the Far East (40%) than it is in Africa (14%) (Kew, 1992). The disparity in HBV e antigen carrier rates in the two regions may be attributed to infection with different HBV genotypes or subgenotypes of the virus (Kramvis et al., 2005a).

The introduction of HBV vaccines in 1981 and their incorporation into national vaccination programmes in 164 countries (WHO, 2008), specifically targeting infants and children, has already made an impact on HBV-related disease burden. A decrease in the incidence of HBV-related HCC has been demonstrated amongst children, adolescents and young adults in Taiwan (Chang et al., 1997; Ni et al., 2007) and a decrease of HBsAg carrier rate in children younger than five years has also been demonstrated in South Africa (Tsebe et al., 2001).

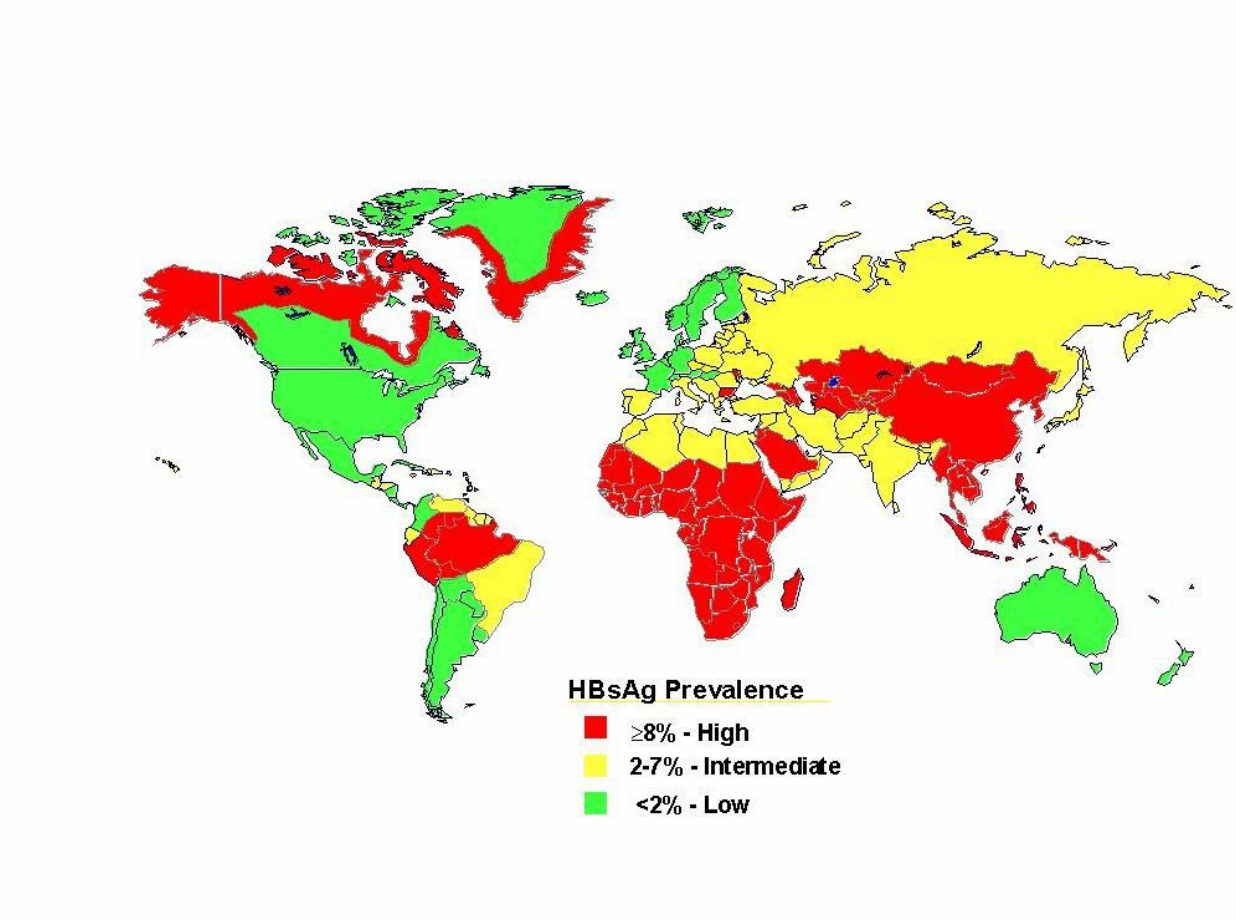


Figure 1.1. Geographical distribution of chronic hepatitis B virus infection (http: and www.dshs.state.tx.us/idcu/disease/hepatitis/hepatitis_b/faqs/, 2005).

1.1. THE FAMILY *HEPADNAVIRIDAE*

The *Hepadnaviridae* is a family of hepatotropic DNA viruses consisting of two genera: *Orthohepadnavirus*, infecting mammals and *Avihepadnavirus*, infecting birds. HBV is the prototype of the mammalian hepadnaviruses. Related viruses have been identified in woodchucks (*Marmota monax*), ground squirrels (*Spermophilus beecheyi*), Arctic squirrels (*Spermophylus parryi kennicotti*) and Woolley monkeys (*Lagothrix logathricha*) and have been named WHV, GSHV, ASHV and WMHBV, respectively (Summers et al., 1978; Marion et al., 1980; Testut et al., 1996; Lanford et al., 1998). Three members of the *Orthohepadnaviruses,* namely, HBV, WHV, and GSHV, are able to induce HCC in their respective hosts.

1.2. VIRAL STRUCTURE

Three types of HBV particles circulate in the blood: Dane particles (virions), spheres and filaments (Dane et al., 1970; Ganem and Varmus, 1987). The 42 nm Dane particles are infectious, their double-shell consists of an outer envelope and inner icosahedral nucleocapsid (figure 1.2A). The envelope consists of cocarboxyterminal large (LHBs), middle (MHBs) and small (SHBs) surface proteins (figure 1.2A). The nucleocapsid is made up of core proteins (HBcAg) that encompass the viral DNA (Gerber and Thung, 1985; Ganem and Varmus, 1987; Seeger et al., 1991; Seeger and Mason, 2000). The viral DNA is associated with at least two proteins: an endogenous DNA polymerase and a protein kinase (PK) (Robinson and Greenman, 1974; Ganem and Varmus, 1987; Gerber and Thung, 1985). The spherical particles have a diameter of 22 nm and the filamentous particles have the same diameter but a variable length. Both particles are non-infectious and contain less LHBs compared with virions (figure 1.2 B & C).

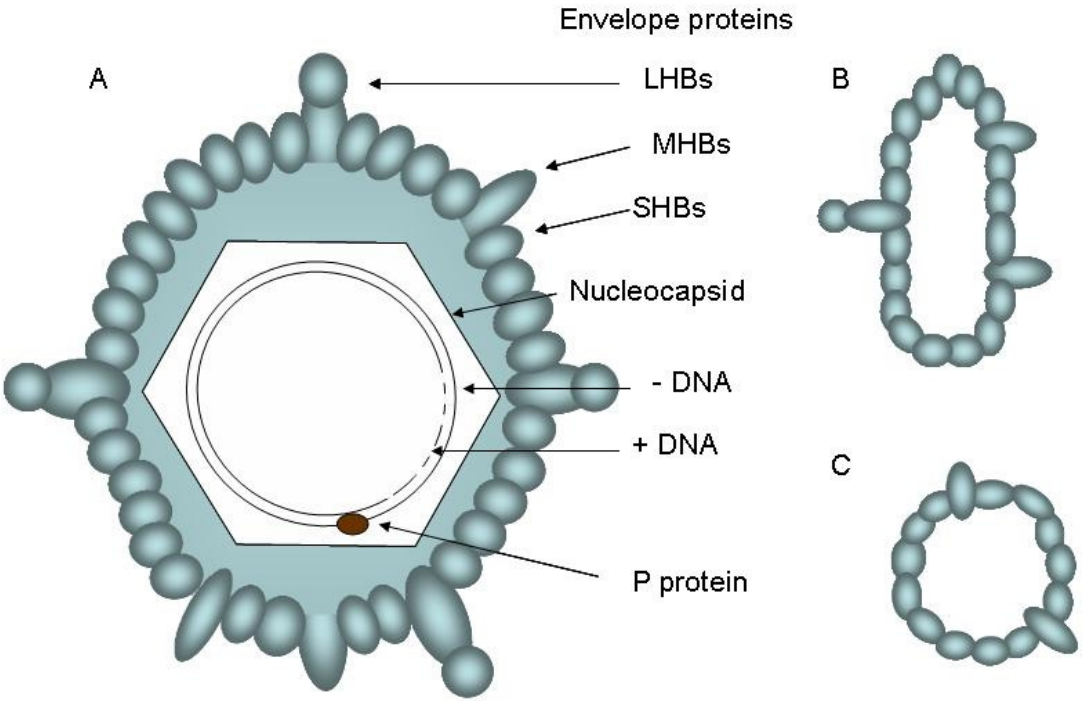


Figure 1.2. A. Diagrammatic representation of a Dane particle or complete HBV virion. The envelope is made up of LHBs, MHBs and SHBs.

B. Filamentous sub-viral particle, which is composed of SHBs and MHBs. C. Spherical sub-viral particle is composed mainly of SHBs and MHBs.

1.3. VIRAL GENOME

The HBV genome size is genotype dependent and varies between 3 182 and 3 248 nucleotides, making it the smallest known animal virus (Feitelson, 1994). The HBV genome within infectious virions is a relaxed circular, partially double stranded DNA molecule (RC-DNA) (Gerber and Thung, 1985; Tiollais et al., 1985; Ganem and Varmus, 1987). This structure is maintained by base pairing between the 5' ends of both strands. The minus strand is complete (3.2 kb), whereas the plus strand is incomplete and varies in length (~1.6-2.8 kb). The 5' end of the minus strand is covalently attached to a terminal protein of the viral polymerase, whereas the 5' end of the plus strand is capped by a RNA oligonucleotide derived from the pregenomic RNA (Gerber and Thung, 1985; Ganem and Varmus, 1987; Beck and Nassal, 2007). The minus strand is the template strand and has an opposite polarity to the viral mRNA transcripts. The genome consists of four overlapping open reading frames (ORFs): the surface (*S*), core (*C*), polymerase (*P*) and *X*. The S ORF encodes the three envelope proteins the LHBs (large), MHBs (middle) and SHBs (small). The *C* encodes for the core protein (HBcAg) and e antigen (HBeAg). *P* encodes the viral polymerase (reverse transcriptase). *X* encodes the HBx transcriptional *trans*-activator (figure 1.3).

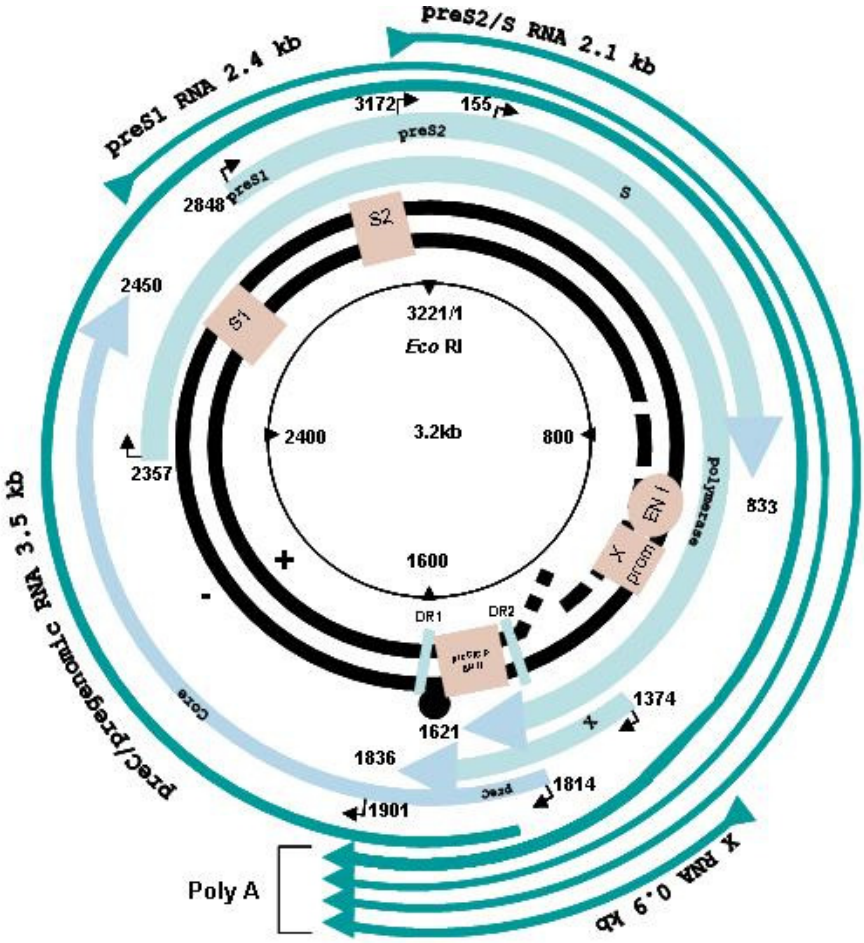


Figure1.3. Organization of the HBV genome. Numbering is from the *Eco*RI cleavage site (Galibert et al., 1979). The outer complete circle represents the minus DNA strand to which the viral polymerase attaches at the 5' end. The inner partially incomplete plus DNA strand is capped with an RNA oligomer at its 5' end. The positions of the viral *cis*-regulatory elements are depicted in the pink circles or boxes. S1 is the pre-S1 gene promoter, S2 is the S gene promoter, X prom is the X gene promoter, preC/CP/EN I is the precore/core promoter and enhancer I and EN II is the enhancer II. The four open reading frames are represented by the thick turquoise green circular arrows (Tiollais et al., 1985) and the HBV transcripts with co-terminal polyadenylation signals are represented by the thin turquoise green circular arrows.

1.4. HBV TRANSCRIPTS

The RC-DNA has to be converted into covalently closed circular DNA (ccc DNA) within the nucleus and this involves removal of the various 5' end appendages, completion of plus strand synthesis and covalent ligation of both strands (Beck and Nassal, 2007). The ccc DNA molecule is used as a transcription template by the host RNA polymerase II to produce the viral transcripts (Beck and Nassal, 2007). The host transcription machinery initiates transcription at 4 different initiation sites on the genome and terminates transcription at a common 3' polyadenylation signal. At least four types of unidirectional RNA transcripts are generated and used to translate seven protein products (Table 1.1).

Table1.1. HBV Transcripts

Transcript	Viral Protein	Start site[1]	Calculated Size[2]
Pregenomic RNA/ Core mRNA	HBcAg & polymerase/RT	1818	3.3
Precore mRNA	HBeAg	1783/1784 or 1790±1	3.5
Pre-S 1 mRNA	LHBs	2807	2.3
Pre-S 2 mRNA	MHBs	3157/3158	2.0
S mRNA	SHBs/HBsAg	3174/3175 5/6 113	1.95
X mRNA	HBx	1226/1242	0.65-0.7

1 Numbering according to (Galibert et al., 1979)

2 Approximate size calculated in kilobases (kb) excluding the poly (dA) tail

Data compiled from: (Cattaneo et al., 1984; Standring et al., 1984; Will et al., 1987; Yaginuma et al., 1987; Kimbi, 2005)

The precore mRNA (3.5 kb) and the core or pregenome RNA (3.3 kb) are functionally different. HBV e antigen (HBeAg) is translated from the precore mRNA template, whereas translation of the core mRNA produces core and polymerase protein. This pregenomic RNA is also reverse transcribed into RC-DNA genomes.

Moreover, the transcription initiation sites for these transcripts occur in close proximity to each other and may vary (Table 1.1). The LHBs is translated from the 2.4 kb mRNA or pre-S1 transcript, and the MHBs and SHBs from the smaller 2.1kb mRNA or pre-S2 transcript. Finally, the HBx protein is produced from the smallest transcript 0.7 kb mRNA (Nassal and Schaller, 1993).

1.5. VIRAL GENE PRODUCTS

At least seven viral proteins are encoded by the four overlapping ORFs, demonstrating the efficient nature of the compact HBV genome (Tiollais et al., 1985). Despite the constraints placed on sequence variation by the ORFs (Mizokami et al., 1997), mutations have been reported in all four ORFs.

1.5.1. SURFACE PROTEINS

The pre-S1 domain consists of 108 or 119 amino acids (aa); this variation is governed by genotypic differences (figure 1.4). The amino (N)-terminus of the pre-S1 domain is primarily involved in infectivity, while the carboxyl (C)-terminus is involved in nucleocapsid envelopment. Initially aa 10-36 in genotype D or 21-47 in genotype A were shown to be essential for binding to hepatocytes (Neurath et al., 1986b). Later, a larger region, aa 3 - 77, was shown to be essential for infectivity (Le Seyec et al., 1999). Further analysis of this domain revealed that aa 9-18 are critical for infection, whereas two accessory domains were also identified (figure 1.4) (Glebe and Urban, 2007). Acylation of glycin 2 with myristic acid is necessary for efficient infectivity (Bruss et al., 1996).The corresponding amino acid positions for genotype A are depicted in figure 1.4. It has been suggested that glycine 2 and glycine 13 are myristoylated in genotype A. Residues 103 - 124 are involved capsid interaction and envelopment (figure 1.4) (Bruss 1997). Overlapping

domains, the cytosolic anchorage determinant (aa 82 - 106) and the Hsc 70 interaction domain (aa 75 - 119), are together implicated in the mechanism involved in the dual topogenic nature of LHBs in the endoplasmic reticulum (ER) membrane or virion envelope (Prange and Streeck, 1995; Loffler-Mary et al., 1997). This dual topogenic nature of the large envelope protein (figure 1.5) is essential for HBV infection, enabling viral attachment to the liver cells using the N-terminal hepatocyte receptor domain of pre-S1 and interaction with nucleocapsids using the C-terminal residues 103-124 (figure 1.4) (Bruss, 1997). In addition to its role in viral infection and sub (viral) particle formation and release, LHBs has also been described as a transcriptional transactivator (Ou and Rutter, 1987; Bruss and Ganem, 1991; Hildt et al., 1996a).

The pre-S2 domain consists of 55 aa. Besides forming part of MHBs and LHBs, the pre-S2 domain contains overlapping B and T cell epitopes and a binding site at aa 120-145 for a neutralizing antibody (Neurath et al., 1984; Chisari and Ferrari, 1995). The pre-S2 also contains a transactivation domain (Hildt et al., 1995; Hildt et al., 2002; Neurath et al., 1984). Phosphorylation of pre-S2 serine 28 by protein kinase C (PKC) leads to constitutive activation of PKC and a tumour promoting pathway (Hildt et al., 1996a). The activation is dependent on the cytosolic orientation of pre-S2, and this is characteristically seen in LHBs and 3' truncated MHBs (MHBs^t) that demonstrate a cytosolic topology (Lauer et al., 1992; Hildt et al., 1996a).

The S ORF encodes three surface proteins: SHBs, is 226 aa contains the S domain and is the most abundant envelope protein and is also known as the HBs antigen (HBsAg); MHBs contains 281 aa and consists of the pre-S2 and S

domains; LHBs is the largest surface protein, containing of 389/400 aa, and consists of the pre-S1, pre-S2 and S domains (figure 1.4) (Ganem and Varmus, 1987; Feitelson, 1994). SHBs and MHBs are usually synthesized in excess in infected hepatocytes or transfected hepatoma cell lines, LHBs constituting less than 10% of the total surface proteins (Huang and Yen, 1993). This ratio is essential for virion assembly and secretion (Huang and Yen, 1993; Wang et al., 2003). LHBs and SHBs, but not MHBs, are critical for virion morphogenesis (Bruss and Ganem, 1991).

The S domain consists of at least three hydrophobic domains occurring between residues 8-22, 80-98 and 170-226 or residues 62-78, 135-153 and 224-281 of MHBs (Tiollais et al., 1985). Two α-helices have been predicted between residues 173-193 and residues 202 and 222 (Persson and Argos, 1994). Residues 23-79 form a loop into the cytoplasm, while residues 100-160 form a secondary structure extending into the ER luminal space or the outer surface of the virion. This region also encompasses the major antigen ‘a’ determinant, the main neutralizing HBV epitope (figure 1.4). Consequently, this antigen, HBsAg forms the basis of the vaccines against HBV (McAleer et al., 1984; Ganem and Varmus, 1987; Feitelson, 1994).

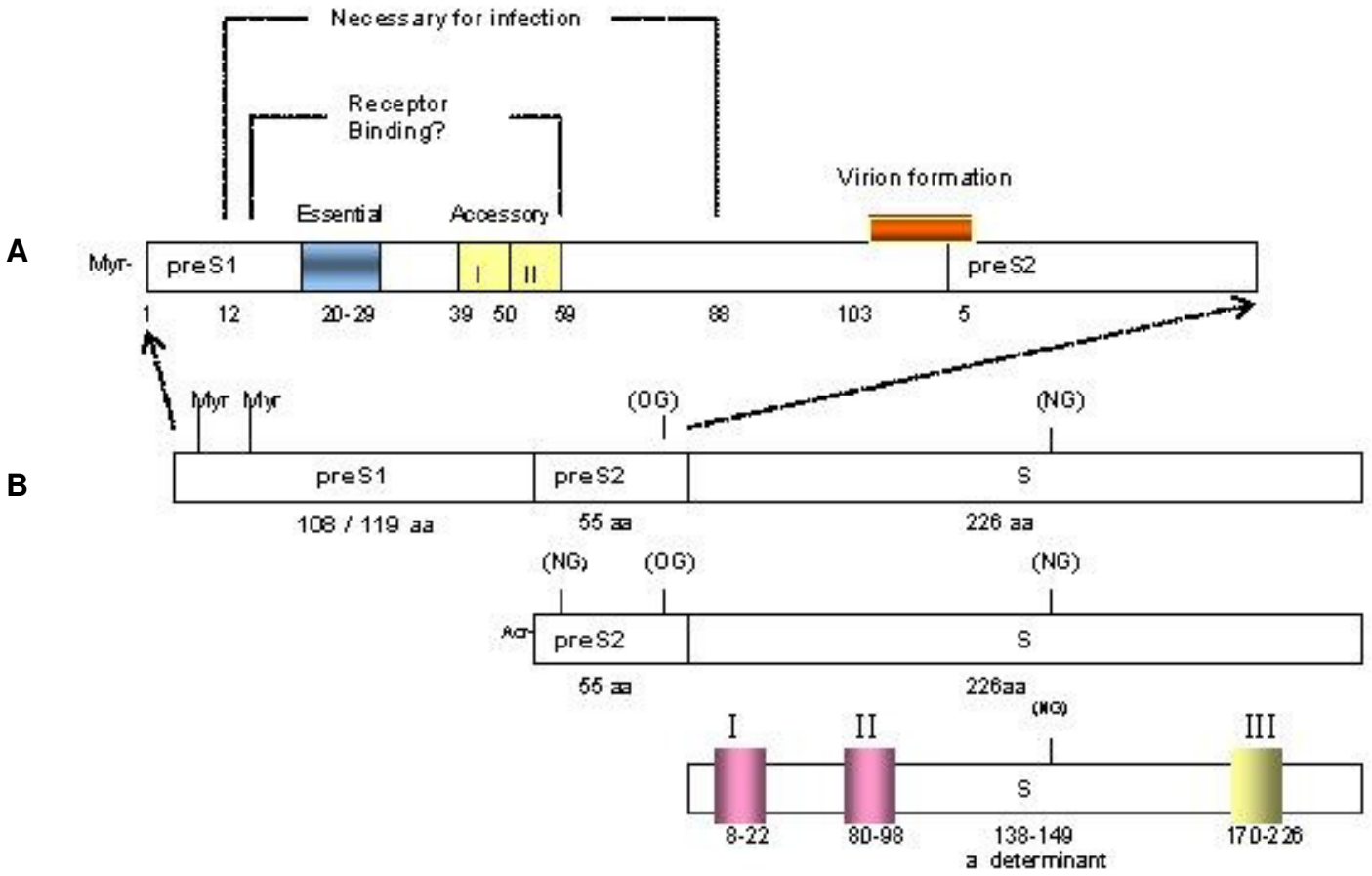


Figure 1.4 A schematic representation of the HBV surface proteins.

Figure 1.4. A schematic representation of the HBV surface protein. A. Pre-S domains essential for virion morphogenesis and viral entry. The putative receptor binding site within pre-S1 is illustrated with the essential domain (aa 20-29) and two accessory domains (aa 39-50 and 50-59). The domain involved in virion formation overlapping the pre-S1/pre-S2 (aa 103-105).
B. Three surface proteins: LHBs, MHBs and SHBs. The pre-S1 domain of the LHB protein is 108 or 119 aa in length depending on the genotype, two putative myristoylation (Myr) sites at glycine 2 and glycine 13, and N-glycosylation (NG) and O-glycosylation (OG) are indicated. The pre-S2 domain of the MHB protein is 55 aa in length and is N-terminally acetylated (Ac). SHBs (HBsAg) are 226 aa in length. Three hydrophobic domains (I, II and III) are indicated as well as the position of the major antigenic 'a' determinant. Adapted from (Glebe and Urban, 2007).

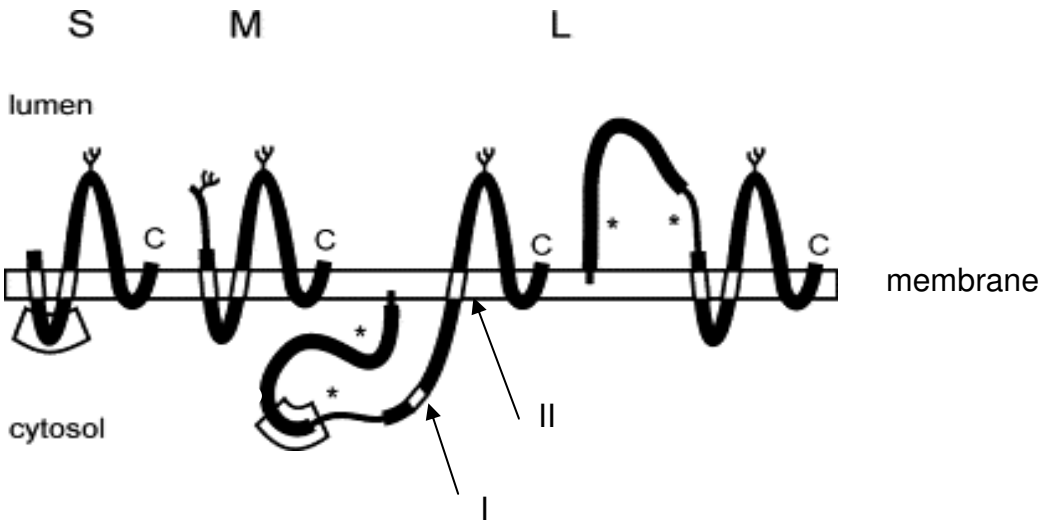


Figure 1.5. Transmembrane topology of the HBV envelope proteins. The transmembrane structure of the SHBs (S) and MHBs (M) is determined by an N-terminal type I (SHBs residues 8-22 or MHBs residues 62-78) and internal type II (SHBs residues 80-98 or MHBs residues 135-153) signal (indicated by arrows). The C-terminal domain (SHBs residues 170-226 or MHBs residues 224-281) is highly hydrophobic and is embedded in the ER lipid bilayer. SHBs and MHBs have a similar topology with the N- and C-termini are folded towards the luminal side. LHBs (L) has the unique ability to adopt two membrane topologies. The cytosolic form allows for interaction with core particles (boxed region) and the luminal form allows for viral interaction with liver cells. *N*-glycosylation sites are shown by small branches on the folded protein. SHBs is *N*-glycosylated at position Asn 146, MHBs is additionally *N*-glycosylated at its Asn residue 4 (Peterson et al., 1982; Stibbe and Gerlich, 1983). The asterisks indicate potential *N*-glycosylation sites. Adapted from (Bruss, 2004).

1.5.2. CORE GENE PRODUCT

The C ORF encodes two proteins, the HBc antigen (HBcAg) and the HBe antigen (HBeAg) (Tiollais et al., 1985). A 21 kDa core protein (HBcAg) is translated from the core mRNA. The protein consists of 183 or 185 aa, determined by viral genotype. The first 147 amino acids are involved in core protein dimerization. Core homodimers form by disulphide bridges between the cysteine residues 61, followed by spontaneous capsid particle formations. Two types of particles form by weak interdimer interactions; those consisting of 90 dimers with a diameter of 30 nm and a triangulation number (T) of 3 and those consisting of 120 dimers and a diameter of 34nm organized with a T of 4 (Crowther et al., 1994). It is unclear which of these occur within the infectious virion. The remaining C-terminal domain is an arginine-rich domain, resembling a protamine and upon capsid assembly this domain is involved in packaging of the pregenome with attached reverse transcriptase (Gallina et al., 1989; Bruss, 2004). The immunodominant region, aa 70 – 90, is presented on the outer surface of the core particle (Salfeld et al., 1989).

The precore mRNA contains an additional initiation codon, 87 nucleotides upstream of the core initiation site. Translation from this in frame AUG, results in the precursor HBe protein of 25 kDa consisting of the pre-C domain of 19 aa followed by a core polypeptide chain (183 or 185 aa). Part of the pre-C region serves as a signal peptide and targets the precursor to the ER membrane, where the N-terminus is cleaved by signal peptidase to yield a 22-23 kDa protein. Additional cleavage of the C-terminus by proteases occurs in the Golgi apparatus resulting in a heterogeneous final mature product of 15-18 kDa known as HBeAg. This protein is secreted into the blood and is also found on the surface of the hepatocytes (Ou et al., 1986; Bruss and Gerlich, 1988; Standring et al., 1988).

1.5.3. POLYMERASE

The P ORF encodes an 830 aa multifunctional protein. The 95 kDa polymerase protein is translated from a core mRNA internal initiation codon by a leaky scanning mechanism (Schlicht et al., 1989; Chang et al., 1990). The polymerase ORF lacks a dedicated promoter and the upstream optimal core start site reduces the efficiency of polymerase expression. These two factors lead to the regulation of translation that probably results in a core to polymerase protein ratio of approximately 240:1, which is sufficient for the HBV life cycle (Hwang and Su, 1998).

The polymerase is made up of four domains (Radziwill et al., 1990; Beck and Nassal, 2007).

- The terminal protein (TP) domain spans amino acids 1-183 at the N - terminus of the protein. Essentially functioning as a primase, tyrosine (Y) 63 is employed for initial primer extension and thus the TP remains covalently linked to the minus strand DNA during replication, eventually preventing phosphorylation of the minus strand DNA (Bartenschlager and Schaller, 1988; Bosch et al., 1988). Not only is this domain involved in priming minus strand synthesis but it is also required for pregenomic RNA encapsidation. TP has no sequence similarity to any other protein in the data base (Ferrer-Orta et al., 2006).
- The spacer region encompasses amino acids 184-354 and serves to separate the N-terminal TP and the reverse transcriptase (RT) (Bosch et al., 1988; Bartenschlager and Schaller, 1988; Radziwill et al., 1990). This region is highly variable and dispensable.

- The RT spans amino acids 355-694. It contains the YMDD motif, which is the catalytic domain that is essential for reverse transcriptase activity.
- The fourth domain, an RNase H domain encompasses amino acids 695-844 in the C-terminus and serves to degrade pgRNA (Radziwill et al., 1990). The RT and RNase H domains together are required as structural components of the pgRNA encapsidation.

1.5.4. HBx PROTEIN

The X ORF encodes a 154 aa protein. HBx is essential for the establishment of *in vivo* HBV infection (Chen et al., 1993). It is a labile protein, found both in the cytoplasm and the nucleus (Henkler et al., 2001; Hoare et al., 2001). This varied cellular localization facilitates its interaction with numerous cellular factors, triggering transactivation pathways. The N-terminus contains a proline serine rich domain, which overlaps with a negative regulatory region (Arii et al., 1992; Poussin et al., 1999). The Kunitz domain functions as the transactivation domain of HBx (Arii et al., 1992; Takada and Koike, 1994). Additional residues essential for transactivation have been described. However, none of the functional domains binds DNA directly (Arii et al., 1992). HBx appears to regulate tumour promoting and proapoptotic pathways and may contribute to hepatocarcinogenesis through these mechanisms (Kim et al., 1991; Koike et al., 1998).

1.6. VIRAL *CIS*-ELEMENTS

1.6.1. PROMOTERS AND ENHANCERS

HBV gene expression is initiated by four independent promoters: Core (pre-C/pregenomic), S1, S2 and X promoters (Moolla et al., 2002). The promoters are

further regulated by two enhancers (ENH I and ENH II) and *cis*-acting negative regulatory elements.

Transcription of both the pre-C and pregenomic mRNAs is directed by the core promoter (nt 1591-nt 1822) (Kramvis and Kew, 1999). The core promoter consists of the basic (basal) core promoter (BCP) and the upstream regulatory region (URR). Although the BCP lacks a distinct TATA-box motif (TATAA), four AT-rich domains AGATTA (nt 1750-nt 1755), TTAAA (nt 1758-nt 1762), TATTA (nt 1771-nt 1775) and CATAAATT (nt 1788-nt 1795) are present (Chen et al., 1995; Yu and Mertz, 1996). Further division into overlapping precore and pregenomic promoters has been described. The precore promoter encompasses nt 1758 to nt 1791, with 1758- TTAAA-1762 as a TATA-like motif and 1788-CATA-1791 as the initiator sequence (Yu and Mertz, 1996). The pregenomic promoter encompasses nt 1788 to nt 1821, with 1788- CATAAAT-1794 as a TATA-like motif and 1817-CAACT-1821 as the initiator sequence (Yu and Mertz, 1996).

The close proximity of the precore and pregenomic promoters requires coordinated regulation and this is achieved by the differential use of response elements by transcription factors within the core promoter. Figure 1.6 shows the distribution of response elements within the BCP and core upstream regulatory sequence (CURS). These include binding sites for ubiquitous transcription factors such as Sp1 and TATA binding protein (TBP) and liver specific transcription factors such as hepatocyte nuclear factor 1 (HNF1), HNF 3 and HNF 4 (Li et al., 1995; Yu and Mertz, 1996; Wang et al., 1998). HNF4 binding site 2 appears to be at the centre of differential transcriptional regulation, which determines the coordinate expression of the pre-C / pregenomic RNA (Kramvis and Kew, 1999).

Several transcription factors bind this site, these include HNF 4, chicken ovalbumin upstream promoter transcription factor (COUP-TF), peroxisome proliferator activated receptor α (PPAR α), retinoid X receptor α (RXR α) and human testicular receptor 2 (TR2) (Raney et al., 1997). HNF4 and TR2 binding to site 2 results in repression of pre C RNA synthesis, the binding of PPAR α/ RXR α heterodimers activates pregenomic RNA synthesis, whereas binding of COUP-TF suppresses pre-C and pregenomic RNA synthesis. It has been suggested that transcription factor binding may block access to respective initiation sequences and abrogates transcription (Kramvis and Kew, 1999).

The URR consists of the negative regulatory element (NRE) and the CURS (Kramvis and Kew, 1999). CURS is a positive regulator of BCP and can cause a 200-2000 fold activation (Yuh et al., 1992). CURS acts in a position and orientation dependent manner. Regulation is affected by cell type and differentiation. CURS is divided into CURS-A and CURS-B, these elements activate the BCP independently or in unison. CURS also contains the α, β, γ and δ domains (figure 1.6) (Yuh et al., 1992). The α, β, γ and δ domains are sequence motifs that act independently within CURS by the interaction with distinct transcription factors (Kramvis and Kew, 1999; Moolla et al., 2002). Overall, the α, γ and δ domains up-regulate the BCP, whereas β down-regulates it (Kramvis and Kew, 1999).

Enhancer II confers liver tissue specificity (Honigwachs et al., 1989; Yuh and Ting, 1993), activates the BCP in orientation-dependent manner and up-regulates the surface and X gene promoters in orientation-independent manner. Enhancer II broadly encompasses nt 1627 to nt 1850 as determined by several studies (Wang et al., 1990; Lopez-Cabrera et al., 1991) and is comprised of two elements: II-A

and II-B. The α box (nt 1646-nt1668) and β box (nt 1704-nt1715) (Yuh and Ting, 1991) occur within elements II-A and II-B respectively. Both elements are essential for EHN II function and these sequence motifs are indispensable.

The NRE functions to suppress the core promoter activity in an orientation independent manner (Yuh and Ting, 1991; Lo and Ting, 1994). It consists of three functional domains: NRE-α, NRE-β and NRE-γ. These three *cis*-elements act in unison to produce a strong inhibitory effect on the core promoter.

The S1 promoter (nt 2710-2800) controls the transcription of the pre-S1 mRNA (2.4 kb). The S2 promoter occurs at nt 2 960-3 180 and directs the transcription of the pre-S2 mRNA (2.1 kb). The CCAAT element within the latter promoter does not only stimulate pre-S2 mRNA production but also has a negative regulatory effect on the S1 promoter, ultimately repressing LHBs production (Lu et al., 1995). The X promoter directs the transcription of the 0.7 kb mRNA, which is translated into the HBx protein.

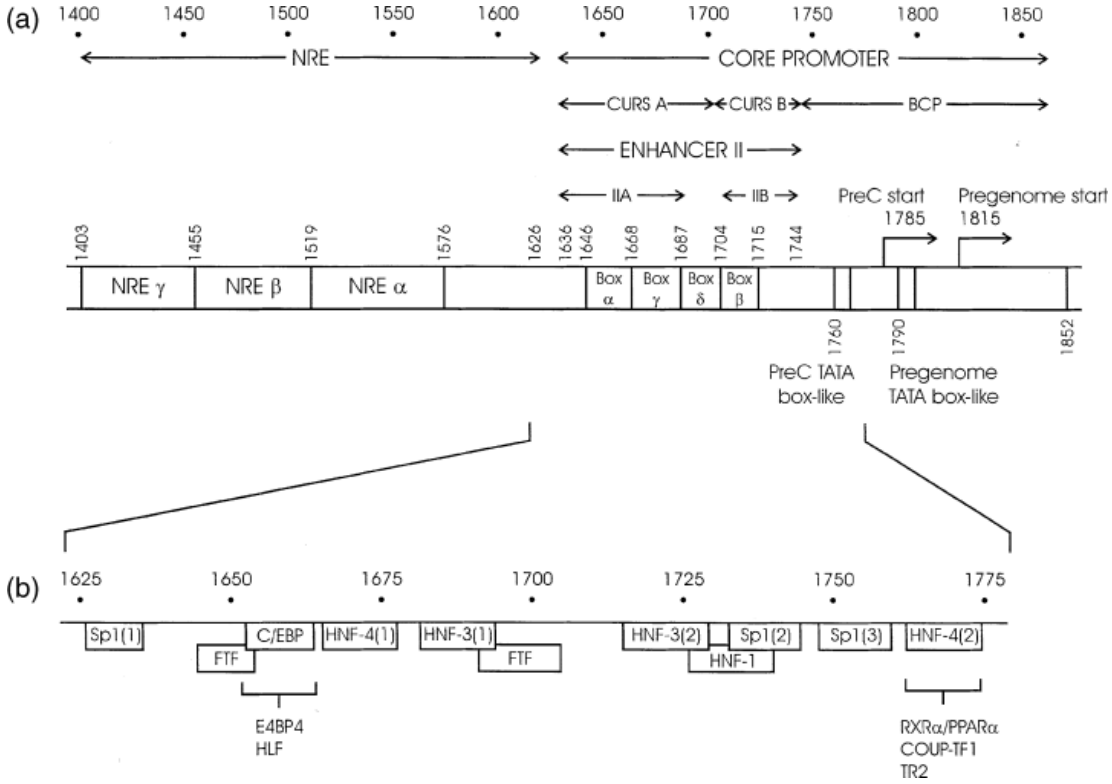


Figure1.6. Core promoter/Enhancer II region (Moolla et al., 2002)

1.6.2. THE ENCAPSIDATION SIGNAL

The 5' end of the pgRNA serves as an encapsidation signal, epsilon (ε) and is highly conserved among the members of the *Hepadnaviridae* (Laskus et al., 1994a; Lok et al., 1994). This 70 nucleotide stem-loop structure consists of two base pairing regions; a 6 nucleotide bulge and a 6 nucleotide loop. This secondary structure is recognized during the packaging of the pgRNA into the capsids. Both full genome transcripts (3.3 kb and 3.5 kb) contain ε. However, only the shorter transcript is encapsidated. In addition to viral encapsidation, ε functions in viral polymerase activation, template restriction, and the initiation of minus strand DNA synthesis. Several mutations within this structure have been shown to affect viral replication and gene expression (Kramvis and Kew, 1998; Guarnieri et al., 2006).

1.6.3. THE POLYADENYLATION SIGNAL

HBV transcription terminates at a common polyadenylation signal (poly A) at position 1916. The TATAAA signal sequence differs from the consensus sequence, and is a weaker signal. HBV replication occurs via an RNA intermediate and this requires that an RNA molecule of more than genome size be produced. During the first passage the RNA polymerase II has to ignore the poly A signal to synthesize an overlength genome. The regulation of RNA polymerase II transcription termination involves the positive and negative effectors of transcription (*pet* and *net*). *Pet* is needed to suppress transcriptional termination during the first transcriptional cycle (Huang and Summers, 1994). *Net* and the poly A signal permit transcriptional termination during the second cycle of transcription (Huang and Summers, 1994; Beckel-Mitchener and Summers, 1997).

1.7. HBV LIFE CYCLE

Host and tissue specificity of hepadnaviruses is determined by N-terminal pre-S1 sequences (Glebe and Urban, 2007). However, additional events subsequent to viral attachment may also play a role. Once inside the hepatocyte, HBV transcription is regulated by liver specific transcription factors, namely HNF-1, HNF-3 and HNF-4. Numerous candidates for the human host receptor for HBV have been investigated without success (Glebe and Urban, 2007). The duck carboxypeptidase (DCPD) has been identified as the receptor for DHBV (Kuroki et al., 1994; Kuroki et al., 1995).

After breaching the cell membrane the capsids are transported to the nucleus. In the nucleus the RC-DNA is released from the nucleocapsid, repaired by host cell machinery to yield ccc DNA (figure 1.7A). The ccc DNA serves as a template for transcription of viral mRNAs, namely: the precore mRNA, pgRNA or core mRNA, pre-S1 mRNA, pre-S2 mRNA, and X mRNA.

ε is not only essential for pgRNA packaging but also initiates viral replication. The viral RT binds to this 3' bulge of this secondary structure and uses Tyr 63 as a primer to synthesize the first 3 nucleotides (5'GAA 3') (Wang and Seeger, 1992). The RT then translocates to the 3' of the pgRNA where the 5'GAA 3' anneals to complementary sequences within DR1 (figure 1.7B). Elongation of the minus strand progresses with the simultaneous degradation of the pgRNA by the RNase H RT activity, sparing only the last 17 to 18 bases that includes a DR1 motif. The short RNA molecule serves as a primer upon translocation to DR2 (Nassal and Schaller, 1996). When the plus strand reaches the 5' end of the minus strand, a

third and final translocation occurs, possibly facilitated by the 9 base terminal redundancy. Viral genome circularization is achieved.

A

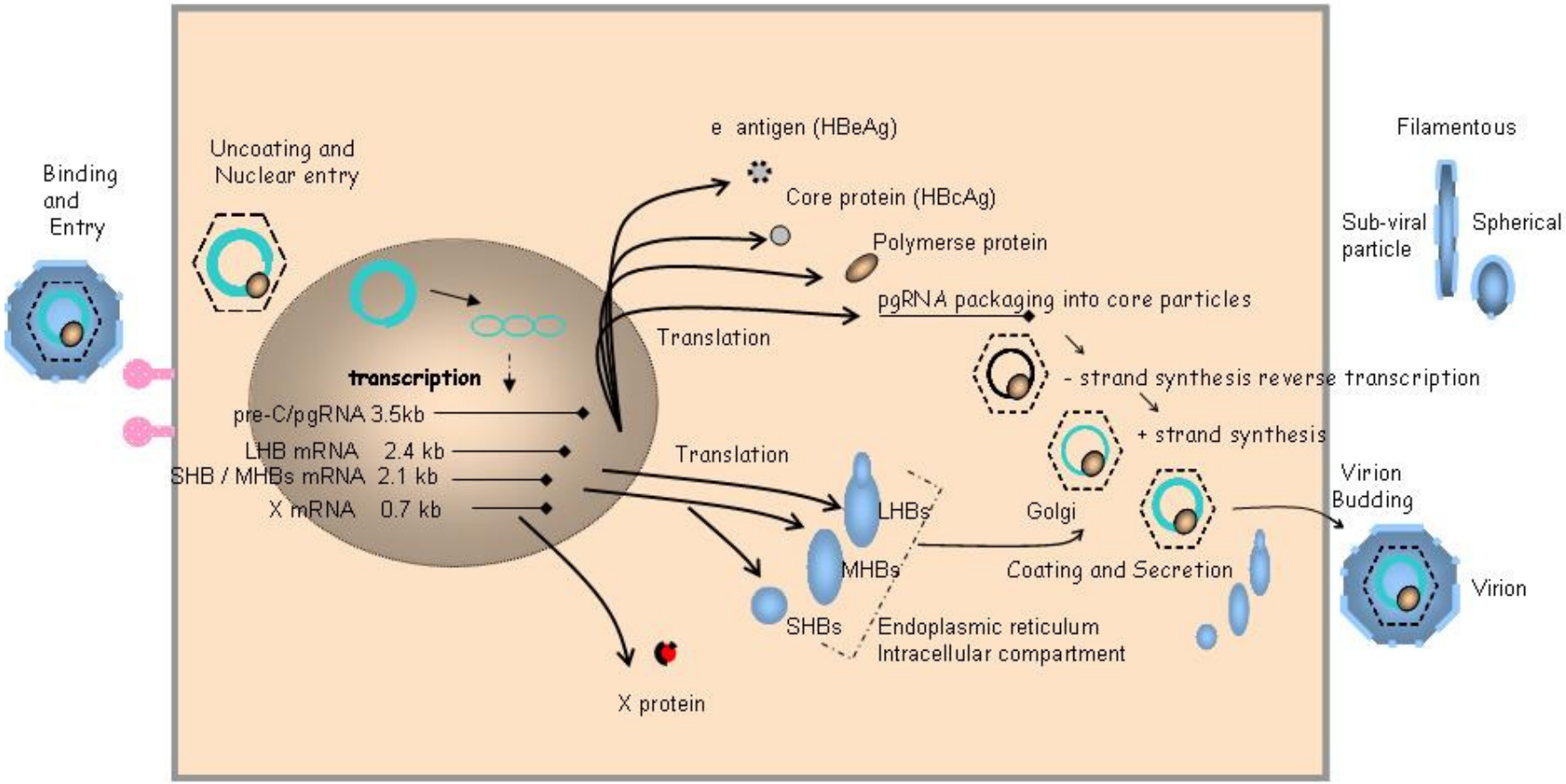


Figure 1.7. The HBV life cycle (A).

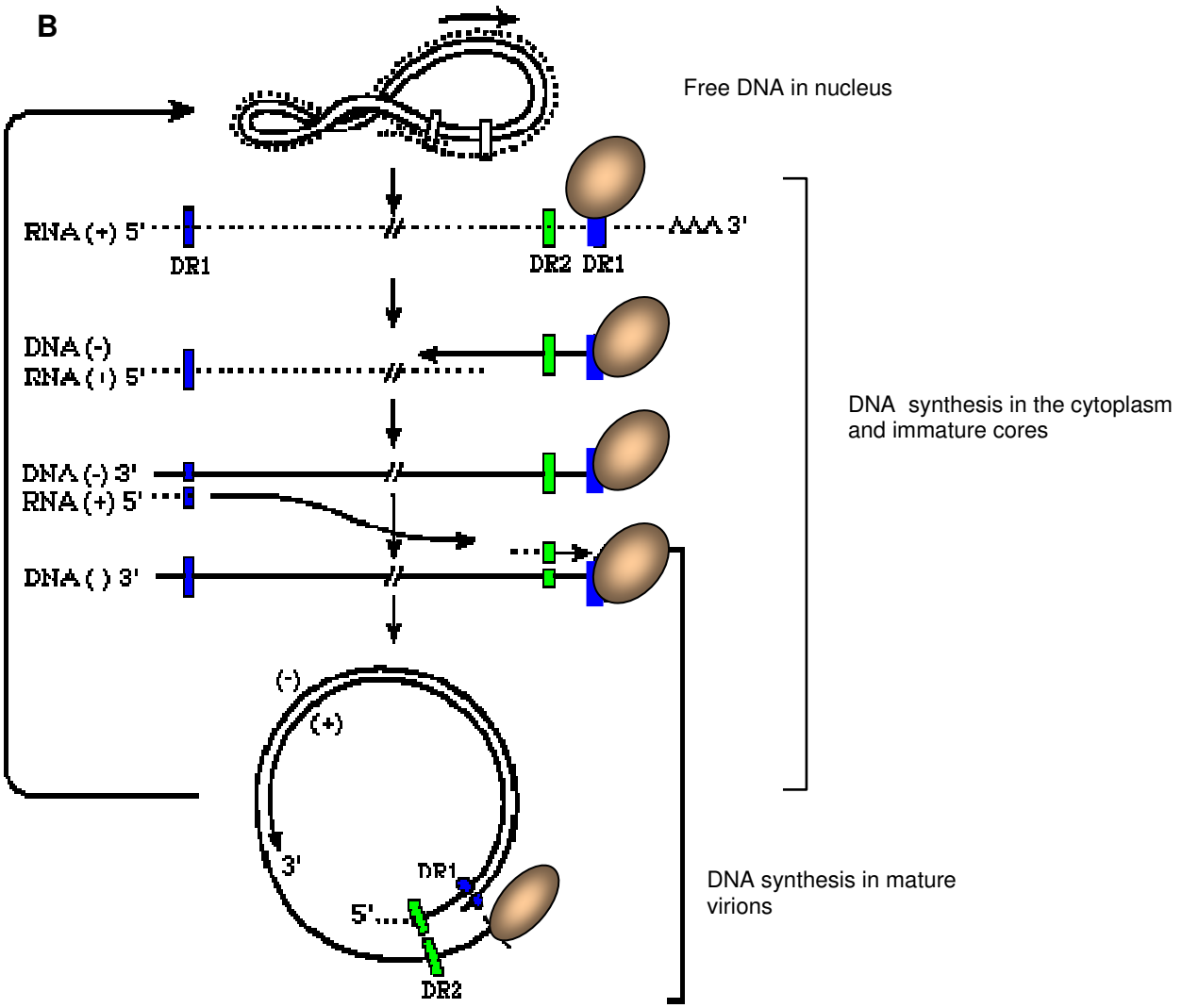


Figure 1.7 A representation of viral replication (B).

Figure1.7. The HBV life cycle. (A). The virion attaches to the hepatocyte via cellular receptors. This HBV specific receptor has not yet been identified. The envelope proteins uncoat in the cytoplasm and the nucleocapsid is transferred to the nucleus. Here, partially double stranded viral DNA is converted to covalently closed circular DNA (ccc DNA). This serves as a template for viral transcription and viral transcripts are translated in the cytoplasm. The 3.5 kb pregenomic mRNA serves as a template for the synthesis of HBcAg (core protein) and viral polymerase/reverse transcriptase (pol/RT). The viral pol/RT and pregenomic RNA are packaged in the core particles. Nucleocapsid maturation occurs once the reverse transcription of the pregenomic RNA has occurred (Summers and Mason, 1982). These mature particles can be shuttled back to the nucleus increasing the cccDNA pool or be enveloped and secreted. (B). A representation of viral replication occurring within the maturing nucleocapsid. Reverse transcription relies upon the DR1 and DR2 sequences indicated in blue and green (Summers and Mason, 1982; hppt and www.stanford.edu/group/virus/1999/tchang/replication.htm, 2007).

1.8. IMMUNO-PATHOGENESIS OF HBV INFECTION

1.8.1. TRANSMISSION

Percutaneous and mucosal exposure to infectious blood and other body fluids are primarily involved in HBV transmission. Transmission is facilitated by numerous forms of human contact, including sexual, mother to child (perinatal), needle sharing, needle stick injuries (health care related) and household (non-sexually) (Francis et al., 1981). In endemic and hyperendemic regions, two major routes of primary transmission occur: perinatally and horizontally during early childhood. Horizontal transmission probably involves the exposure of lesions or scratches to contaminated blood and body fluids from close family (household) members and playmates (Francis et al., 1981) . In comparison, infection in areas with low endemnicity has been associated with certain high risk sexual behaviour. Sex with multiple partners in heterosexual and homosexual individuals as well as intravenous drug use with sharing of various drug-related utensils lead to high rates of HBV disease (Shepard et al., 2006).

1.8.2. PATHOGENESIS OF HBV INFECTION

Approximately 90% of HBV infections acquired early in life (neonates, infants or young children) progress to chronic infections (Kew et al., 1987) and these individuals are at a greatly increased risk for developing HCC. Infections acquired during adulthood are usually acute, clinically evident, and self-limiting. Acute infections almost invariably resolve within 6 months (McMahon et al., 1985). Resolution of acute infections requires the inhibition of viral replication and the elimination of infected hepatocytes by a combination of antiviral cytokine activity and cytolysis of HBV infected hepatocytes (Chang and Lewin, 2007). The various clinical manifestations of HBV infection are depicted in figure1.8.

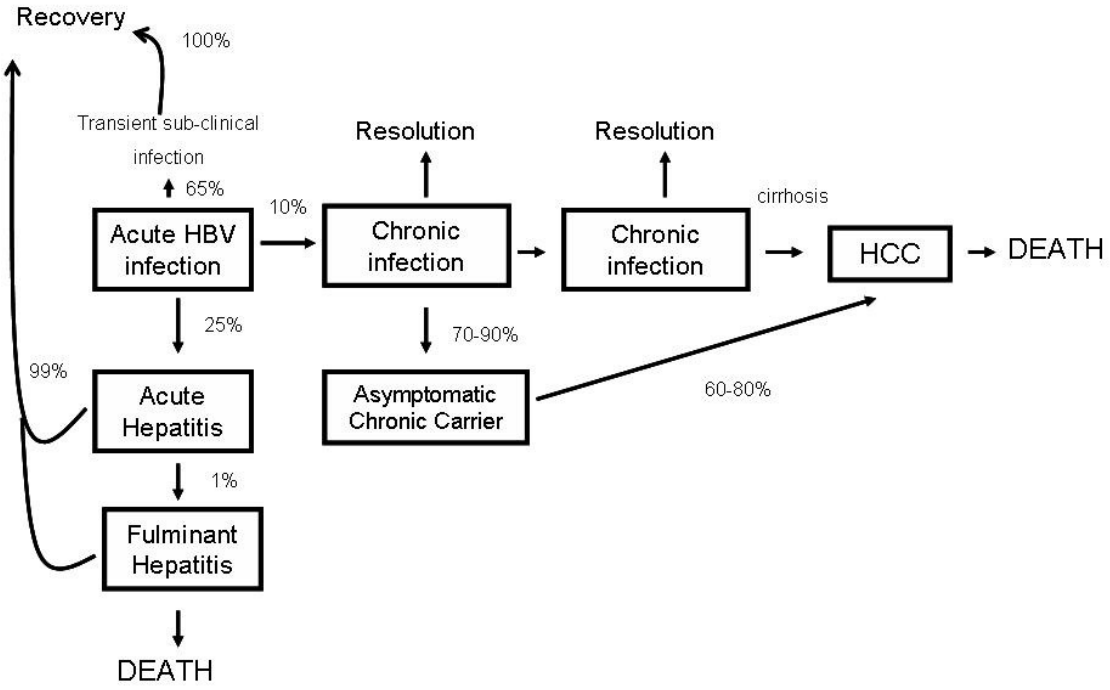


Figure 1.8. Outline of HBV-induced disease and the prevalences of the various clinical sequelae characteristic of southern Africans.

1.8.3. ACUTE HEPATITIS

Ninety five percent of HBV infections in adults are acute and self-limiting (Stevens et al., 1975). Fulminant hepatitis, which involves the unabated necrosis of hepatocytes, occurs in 1% to 2% of acute infections (http://www.who.int) resulting in liver failure and death (Liu et al., 2001).

Subsequent to infection with HBV, infected hepatocytes release IFN-α and IFN-β, which leads to the recruitment of antigen presenting cells (APC). APCs use the major histocompatability complex class I and class II glycoproteins to present viral epitopes to cytotoxic T lymphocytes ($CD8^+$ T-cells) and helper T-cells (Th1 and Th2), respectively. Viral clearance is achieved by a vigorous polyclonal and multi-specific $CD8^+$ T-cell response, primed by Th1 ($CD4^+$T-cells), apposed to a strong

humoral response primed by Th2 $CD4^+$cells. Generally $CD4^+$ T-cells responses are more frequently detected to HBcAg than to other HBV proteins (Ferrari et al., 1991). On the other hand, HB-specific $CD8^+$ T-cells respond to multiple targets, which include epitopes within several proteins: surface (aa 183-191, aa 250-258, aa 335-343), core (aa 18-27) and the polymerase (aa 455-463, aa 575-582) (Maini et al., 2000). In addition, cytokines such as IFN-α, IFN-β and IFN-γ significantly reduce viral replication by various mechanisms. These cytokines are implicated in significantly reducing the high levels of viremia typical of the initial stage of acute hepatitis (Guidotti et al., 1996a).

1.8.4. CHRONIC (PERSISTENT) HEPATITIS

HBV infections acquired early in life or in individuals with a compromised immune system at the time of infection, favour viral persistence (Seeger and Mason, 2000). This is reflected by the high prevalence of viral persistence subsequent to acute infections in neonates (90%) compared with 25-50% in children between the ages of 1 and 5 years and 10% in adults (McMahon et al., 1985). Other high risk groups include immunocompromised individuals such as end-stage renal disease patients, individuals with Down's syndrome, lepromatous leprosy patients and individuals receiving immunosuppressive therapy (Hollinger et al., 1972; Szmuness et al., 1974; Shepard et al., 2006).

Chronic HBV infection is defined by the persistence of HBsAg in the serum for more than 6 months. Asymptomatic chronic carriers (ASC) appear healthy with normal serum aminotransferase levels. The diagnosis of asymptomatic chronic HBV infections may not be straight forward. HBsAg, HBcAg and HBV DNA levels are variable. HBsAg may or may not be detected by conventional assays as a

result of the presence of HBV DNA integrants or mutations (Owiredu et al., 2001a; Allain, 2004), this could be attributed to the gross rearrangements typical of integrated HBV genomes. HBcAg is detected in 5-50% of all chronic HBV carriers. High HBV DNA levels are reported in 5-10% of asymptomatic HBV carriers (Bonino et al., 1986).

HBeAg appears to be instrumental in inducing immune tolerance and HBV persistence after perinatal infection (Milich et al., 1990; Milich et al., 1998; Chen et al., 2005). Infants born to HBeAg positive mothers often develop persistent HBV infections. Using an HBeAg-expressing transgenic mouse model, it has been demonstrated that these mice achieved tolerance to HBc/HBe T-cell determinants, do not produce antibodies to HBeAg, but do produce antibodies to HBcAg (Milich et al., 1990). These characteristics mimic the immunology of infants born to HBeAg positive mothers. HBeAg blocks antiviral cytokines (Whitten et al., 1991). HBeAg may contribute to the depletion of HBeAg- and HBcAg-specific Th1 CD4+ cells through the death receptor (Fas) pathway, and in this way regulate the host immune response (Milich et al., 1998). Additionally, HBeAg may promote the production of anti-inflammatory cytokines, ultimately reducing the CTL response (Milich et al., 1998). Viral clearance coincides with anti-HBs or anti-HBe seroconversion and occurs in ~2% and 10% of chronic carriers annually.

Chronic carriers have a 90% probability of remaining HBV positive for the remainder of their lives, and are at an increased risk for developing HCC compared with uninfected individuals (Beasley et al., 1981).

The current therapeutic approach for chronic HBV infection addresses both viral replication and the host immune response. This includes treatment with IFN-α or nucleoside analogues such as lamivudine, entecavir or adefovir (Dieterich, 2007).

1.8.5. HEPATOCELLULAR CARCINOMA

HCC is the fifth most common malignant disease worldwide and the most common primary tumour of the liver (Parkin, 2001). Chronic HBV infection is the cause of 70 - 80% of all HCC cases. A case/control study conducted in South Africa revealed that HBsAg positive carriers have a 9-33 times increased risk of developing HCC (Kew, 2002). Furthermore, the global distribution of chronic HBV infection and HCC incidence clearly coincide (figure 1.1). The incidence of HCC is particularly high in East and South East Asia and sub-Saharan Africa, with HBV carrier rates between 10-15%. Intermediate incidences occur in western Asia, Central America, eastern and southern Europe with carrier rates between 2-7%. The remaining countries have low carrier rates (<2%) and low incidences of HCC (Parkin et al., 1999; Parkin, 2001).

Some of the highest incidences of liver cancer have been reported in the Black populations of central Africa (27.8 per 100 000 of the population per annum amongst men and 13.4 per 100 000 of the population per annum amongst women) (Parkin et al., 2008), 21.1 and 8.6 per 100 000 per annum in East Africa amongst men and women respectively (Parkin et al., 2008) and 7.0 and 2.5 per annum in southern Africa amongst men and women respectively (Parkin et al., 2008). An estimated 50 000 new cases of HCC are diagnosed in Africa each year (Kew, 2006).

Patients with HCC, whether in regions with high, intermediate, or low incidence of the tumour, invariably have a significantly higher prevalence of the markers of current HBV infection in serum and liver tissue than matched controls (Szmuness, 1978; Kew, 1981). This has been demonstrated in sub-Saharan Black Africans where current markers for HBV infection have been detected in the serum of 71% of HCC patients in southern Africa (Kew et al., 1979), 63% in Senegal (Kew et al., 1987) and 61% in The Gambia (Kirk et al., 2004). Most Black Africans that are positive for serum HBsAg also express the antigen in the liver tissue (Kew et al., 1980).

The prognosis of black African patients with HCC is worse compared with patients with HCC in industrialized countries, with a survival time as low as six weeks from the time of diagnosis (Kew, 2006). The most common cause of death in black African patients with HCC is advanced malignant disease (Kew and Paterson, 1985). The coexistence of cirrhosis has little impact on the prognosis (Kew, 1989).

Chronic HCV infection, cirrhosis and environmental exposure to aflatoxin B (*Aspergillus flavus*) and dietary iron overload have also been implicated in the aetiology of HCC in Africa (Asare et al., 2006).

1.9. HBV GENOTYPES: MOLECULAR CHARACTERISTICS AND GLOBAL DISTRIBUTION

HBV can be classified into eight genotypes (A to H) based on a complete genome nucleotide divergence of 7.5%-8% (Norder et al., 1993; Stuyver et al., 2000; Kramvis et al., 2008) and >4% at the S gene (Norder et al., 1992). These genotypes also differ in size as a result of variations present in the core or pre-S1 regions (Table 1.2). HBV has also been classified into ten serological subtypes

ayw1, ayw2, ayw3, awy4, ayr,adw2, adw3, adw4, adrq+ and *adrq-* (Norder et al., 1992; Magnius and Norder, 1995; Arauz-Ruiz et al., 2002). These are based on the amino acid variations in the S ORF. The major antigenic domain is the common 'a' determinant, two alleles are defined by Thr and Ile at position 126. Two mutually exclusive sub-determinants occur at positions 122 and 160 (Okamoto et al., 1987b) specifying *d/y* and *w/r* (Le Bouvier, 1971; Bancroft et al., 1972). Further variation at residues 122, 127 and 160 has led to the classification of the additional serological subtypes. Genotypes and serotypes do not strictly correspond (Kramvis et al., 2008) (Table 1.2) and serological subtypes do not always reflect genetic relatedness (Kramvis et al., 2005a).

Table 1.2 Genotype Variations

Genotype	Size	Sequence Variation	Serotype
A	3 221	Insertion of aa 153 & aa 154 in the core protein	*adw2, ayw1, ayw2 & adw4*
B	3 215		*adw2, ayw1 & adr*
C	3 215		*adw , adw 2, adwr, ayw2, ayw3, adrq+, adrq- & ayr*
D	3 182	Deletion of aa 1-11 in pre-S1	*ayw2, ayw3, ayw4 & adw3*
E	3 212	Deletion of aa 11 in pre-S1	*ayw4*
F	3 215		*adw4q-, adw2, adw4 & ayw4*
G	3 248	Insertion of 12 amino acids in the core protein Deletion of aa 11 in pre-S1	*adw2*
H	3 215		*adw4*

Adapted from (Kramvis et al., 2005a; Schaefer, 2007a)

The global distribution varies between genotypes (Table 1.3). Genotype A and D appear to have the widest distribution. Genotype A is prevalent in southern Africa, central Africa, Brazil, Northwestern Europe and the USA. Genotype B and C are concentrated in the Far East and Asia. Genotype D also occurs in southern Africa, North Africa, the Middle East, India, but prevails in the Mediterranean. Genotype E is concentrated in West Africa. Genotype F has been identified in Polynesia, south and Central America. The distribution of genotype G has not been clearly defined yet, as only specific isolates have been detected in the USA, the UK, France and Germany. Although genotype A predominates in southern Africa it co-circulates with genotype D in this region.

HBV genotypes are divided into subgenotypes based on less than 4% nucleotide difference (Kramvis et al., 2005a). Subgenotypes have been identified in genotypes A, B, C, D and F (Table1.3).

Table 1.3 Geographical Distribution Genotypes/Subgenotypes[1]

Genotype	Subgenotype	Alternative nomenclature	Geographic Origin	Reference
A	A1	Aa, A'	Africa, (Asia, South America)	(Kimbi et al., 2004; Kramvis et al., 2002)
	A2	A, A-A'	Europe	(Sugauchi et al., 2004a)
	A3	Ac	Cameroon, Gabon	(Kurbanov et al., 2005; Makuwa et al., 2006)
	(A4)		Mali	(Olinger et al., 2006)
	(A5)		Nigeria	(Olinger et al., 2006)
B	B1	Bj	Japan	(Sugauchi et al., 2004b; Sugauchi et al., 2002)Sugauchi et al. 2002; Sugauchi et al. 2003
	B2	Ba	Asia without Japan	
	B3		Indonesia, Philippines	(Norder et al., 2004)
	B4		Vietnam	(Norder et al., 2004)
	B5		Philippines	(Nagasaki et al., 2006; Sakamoto et al., 2006)
	B6		Canadian Inuit (Eskimo)	(Kramvis et al., 2008)
C	C1	C2, Cs	South East China (Vietnam, Myanmar, Thailand, Southern China)	(Chan et al., 2005; Tanaka et al., 2005; Huy et al., 2004)
	C2	C1, Ce	Far East (Korea, Japan, Northern China)	
	C3		Micronesia	(Norder et al., 2004)
	C4		Australia	(Sugauchi et al., 2001)
	C5		Philippines	(Sakamoto et al., 2006);
D	D1		Middle East, Mongolia, Belarus, Europe?	
	D2		Germany, India?	
	D3		Alaska, Sweden, France and South Africa	(Kimbi et al., 2004; Banerjee et al., 2006)
	D4		Somalia, Oceania (Australia Sagauchi et al. 2001)	(Sugauchi et al., 2001)
	D5		East India	(Banerjee et al., 2006)
E	-	-	West Africa	(Odemuyiwa et al., 2001)
F	F1		South & Central America	(Kato et al., 2005; Norder et al., 2003)
	F2		South America	(Naumann et al., 1993)
	F3		Bolivia	(Devesa et al., 2004; Huy et al., 2006)
	F4		Argentina	(Devesa et al., 2004; Huy et al., 2006)
G	-	-	USA, UK, Germany and France	(Bhat et al., 1990; Tran et al., 1991; Stuyver et al., 2000)
H	-	-	Central America	(Arauz-Ruiz et al., 2002)

Adapted from (Schaefer, 2007b)

1.9.1. GENOTYPE A

Genotype A has a 6 nucleotide insert at the carboxyl terminus of the core gene and results in an additional two amino acids at positions 153 and 154, which are not found in the other genotypes. Genotype A can be phylogenetically divided into at least five subgenotypes A1-A5 (Table 1.3). Subgenotype A1 isolates were initially described in South Africa (Bowyer et al., 1997), Malawi (Sugauchi et al., 2003b) and later in Philippines, India, Nepal, Bangladesh and Brazil (Sugauchi et al., 2004a). Subgenotype A1 has unique characteristics in all the ORFs and in the transcriptional regulatory elements (Kimbi et al., 2004). The differences may contribute to the distinct virological characteristics of subgenotype A1 (Kramvis and Kew, 2007b), specifically the manner in which HBeAg is cleared from the serum and the lower viral loads characteristic of subgenotype A1 infections (Tanaka et al., 2004). Certain amino acid changes in pre-S1 region were shown to be distinctive of subgenotype A1 (Kimbi et al., 2004). Subgenotype A2 is found in Europe (Sugauchi et al., 2004a), whereas subgenotype A3 isolates have been identified in Cameroon and Gabon (Kurbanov et al., 2005; Makuwa et al., 2006). Recently, subgenotype A4 and A5 isolates have been described in Mali and Nigeria, respectively (Olinger et al., 2006).

1.9.2. GENOTYPE B

To date, six subgenotypes (B1-B6) have been described for genotype B (Table 1.3) (Sugauchi et al., 2002; Norder et al., 2004; Sugauchi et al., 2004b; Nagasaki et al., 2006; Sakamoto et al., 2006; Kramvis et al., 2008). Subgenotype B1 is concentrated in Japan and is considered to be the authentic genotype B. In comparison subgenotype B2 predominates in Asia (excluding Japan), and is essentially genotype B with a genotype C precore region with nucleotide

boundaries between 1740 -1838 and 2443-2485, defining this subgenotype as a B/C recombinant.

1.9.3. GENOTYPE C

Genotype C has been classified in five subgenotypes C1-C5 (Sugauchi et al., 2001; Huy et al., 2004; Chan et al., 2005; Tanaka et al., 2005; Sakamoto et al., 2006). These subgenotypes also have different regional distribution patterns (Table 1.3).

1.9.4 GENOTYPE D

The most recent data reveal that genotype D also has five subgenotypes D1-D5, however their exact distribution pattern is not clear (Table 1.3) (Sugauchi et al., 2001; Kimbi et al., 2004; Banerjee et al., 2005). Genotype D isolates typically have a length of 3 182 nucleotides and have a signature 33 bp deletion at the beginning of the pre S1 gene. This results in the deletion of the first 11 aa of the pre-S1 region.

1.9.5. GENOTYPE E

As a result of the lack of separation between genotype E and genotype D in the C and X ORFs (Bowyer and Sim, 2000), genotype E appears to be closely related to genotype D. A deletion of amino acid 11 (3 nt deletion) in the pre-S1 domain is characteristic of this genotype (Norder et al., 1994). Another distinguishing characteristic of genotype E is the introduction of an additional initiation codon upstream of the pre-S2 start site, caused by a 3095 A→ G substitution (Moolla et al., 2002). Predictably, this may result in the expression of an elongated MHBs protein consisting of an additional 36 aa (317aa in length) (Kramvis et al., 2005b). The absence of subgenotypes in genotype E and the low intragroup divergence

suggests that this genotype has been recently introduced into the human population (Mulders et al., 2004).

1.9.6. GENOTYPE F

Genotype F may represent the first split from the human hepadnaviral ancestor, because it diverges from the other genotypes by 14% (Norder et al., 1994). Recent data has shown that genotype F can now be divided into four subgenotypes F1-F4 (Naumann et al., 1993; Norder et al., 2003; Devesa et al., 2004; Kato et al., 2005; Huy et al., 2006). The first two subgenotypes are distinguishable by a nucleotide substitution at position 1858 in the precore region and an amino change at position 45 of the S gene. Subgenotype F1 isolates are mainly from Central America and have a T at position 1858 and Thr at residue 45 of HBsAg. Subgenotype F2 isolates are mainly from South America and Polynesia and possess a C at position 1858 and Leu at residue 45 of HBsAg (Norder et al., 2003). Subgenotype F3 and subgenotype F4 have been identified in Bolivia and Argentina, respectively (Devesa et al., 2004; Huy et al., 2006).

1.9.7. GENOTYPE G

All genotype G isolates have length of 3 248 nt and are characterised by a 36 bp insertion within the core gene at nucleotide position 1905 (Bhat et al., 1990; Tran et al., 1991). Similar to genotype E, genotype G has a 3 bp deletion in the amino terminal end of the pre-S1. Another distinctive characteristic of this genotype is the absence of HBeAg expression caused by the presence of two translational stop codons at positions 2 and 28 of the precore region (Kato et al., 2002). However, HBeAg expression is often detected as a result of coinfection with genotype A (Kato et al., 2002).

1.9.8. GENOTYPE H

After full genome analysis of isolates from Nicaragua and then Mexico the existence of genotype H was confirmed. Owing to its close similarity to genotype F, it has been speculated that genotype H split for genotype F within the New World (Arauz-Ruiz et al., 2002). Genotype H also possesses distinctive mutations within the S domain at Val 44 and Pro 45.

1.10. HBV SEQUENCE HETEROGENEITY

It is becoming more evident that HBV sequence variation can affect the fundamental aspects of the viral life cycle, which in turn may influence the course and severity of HBV infection and disease. Since different mutations are more prevalent in certain HBV genotypes and genotypes have different geographical patterns, perhaps different mutation patterns account for disease in different regions of the world.

Mutations are common in HBV and have been reported throughout the genome including essential regulatory elements. HBV has an unusually high mutation rate (1.4×10^{-5} to 5×10^{-5} nucleotide substitutions per site per year) (Okamoto et al., 1987a; Mimms, 1995; Fares and Holmes, 2002; Kramvis et al., 2005a), although sequence variation is constrained by the overlapping and compact nature of the HBV genome (Mizokami et al., 1997). HBV sequence heterogeneity is probably achieved by a variety of mechanisms. The viral reverse transcriptase lacks proofreading activity (Steinhauer and Holland, 1986) and this is probably exacerbated by the high viral turnover during infection. Splicing of the pregenomic RNA creates large deletions, which interrupts several genes (Gunther et al., 1997a; Huang et al., 2000; Soussan et al., 2000). HBV recombination events are

probably mediated by topoisomerase I (Kew et al., 1993; Pourquier et al., 1999) and aberrant viral replication (Sommer et al., 1997). *In vivo* analysis has also identified G to A hypermutations in HBV genomes that may be induce by antiviral cellular cytidine deaminases (Gunther et al., 1997b; Bonvin et al., 2006).

In addition, to these direct biochemical mechanisms, host immune-virus interaction and selective pressure imposed exogenously by vaccination and antiviral treatment, can also lead to HBV sequence heterogeneity. These mechanisms probably contribute to the establishment of stable, replication competent variants in the form of HBV genotypes and subgenotypes or variants that are expressed transiently and/or dominantly during the course of persistent infection.

1.10.1. HBV GENOTYPES AND DISEASE PROGRESSION

Most of the studies examining the impact of HBV genotype on liver disease have been carried out in hyperendemic regions such as Asia and Africa. In this regard the effect of HBV genotypes has largely been neglected in Western countries.

Genotype B and C are the predominant strains in Asia and comparative studies have shown that genotype C causes more severe liver disease than genotype B (Sugauchi et al., 2002; Sugauchi et al., 2003a; Yuen et al., 2004). Early studies in South-east Asia reported a higher prevalence of liver disease and liver dysfunction in genotype C than genotype B infected patients. Studies in the Far East also correlated genotype C with an increased risk of HCC development (Yuen et al., 2004; Orito et al., 2005). Interestingly, different subgenotypes prevail in these regions: subgenotype C1 prevails in South-east Asia and subgenotype C2 in the Far East (Chan et al., 2005; Tanaka et al., 2005).

A high frequency of the 1762 A→T and 1764 G→A (1762T1764A hereafter) BCP mutation has been reported in genotype C, even though this mutation is linked to decreased HBeAg expression, no correlation was evident between the prevalence of the mutation and serum HBeAg levels (Orito et al., 2001; Yuen et al., 2004). There are also conflicting reports regarding HBV DNA levels in infected patients: higher levels have been reported for genotype C (Yu et al., 2005; Du et al., 2007) compared to B in some studies and not in others (Yuen et al., 2004; Yu et al., 2005; Du et al., 2007). The interpretation of this may be subject to the HBeAg status of the patients.

While genotype B may be relatively less aggressive when compared to genotype C, there are also regional differences within this genotype with regard to clinical presentation and outcome of disease (Lindh et al., 1999). Genotype B in Taiwan is associated with the development of HCC at a young age (Kao et al., 2000), while genotype B in Japan is associated with the development of HCC at a significantly older age (Orito et al., 2005). This may be attributed to the prevalence of subgenotype B2(Ba) in the former regions and subgenotype B1(Bj) in the latter (Sugauchi et al., 2002).

Although genotype A, D and E circulate in southern Africa, most of the research has been focused on genotype A and hence a clearer understanding of its role in disease progression has emerged. A case control study carried out in this region used sub-genomic PCR, RFLP analysis and DNA sequencing to determine the prevalent genotype and revealed that patients infected with subgenotype A1 were 4.5 times more likely to develop HCC than those infected with non-A genotypes. In addition, genotype A infected patients were on average 6.5 years younger than

those infected with non-A (P< 0.05) (Kew et al., 2005). Stable traits which are characteristic of subgenotype A1, which incude the 1862T mutation (Tanaka et al., 2004) and substitutions at nts 1809-1812 (Kimbi et al., 2004; Sugauchi et al., 2004a) and additional mutations such as 1762T1764A (Baptista et al., 1999) may contribute to early loss of HBe or anti-HBeAg seroconversion and affect HBV DNA levels, underpinning the differences between subgenotype A1 and non-A genotypes (Baptista et al., 1999; Kramvis and Kew, 2007a).

1.10.2. HBV MUTANTS AND DISEASE PROGRESSION

1.10.2.1. CORE PROMOTER AND ENHANCER II REGION

Transcription of both the pre-C and pregenomic mRNAs is directed by the core promoter (CP) (figure 1.6). The CP consists of the BCP and the URR. Two highly conserved regions, nts 1770-1808 and nts 1813-1849, have been identified within the CP (Baptista et al., 1999). Several mutations generally cluster between nt 1750 and 1780. These include point mutations, deletions and insertions.

Subgenomic analysis of BCP region (nt1661 to nt1921) revealed a significantly higher prevalence of the 1762T1764A double mutation in black African HCC patients in comparison to matched asymptomatic carriers (66% vs. 11%); however this did not correlate with HBeAg status (Baptista et al., 1999). *In vitro* analysis of the 1762T1764A in genotype A revealed an enhanced effect on viral replication and reduced HBeAg expression (Guarnieri et al., 2006). In addition to implicating these specific mutations in HCC, which is supported by studies done elsewhere in the world (Hsia et al., 1996) the South African study also revealed additional changes in the BCP. The 1753C mutation was present in ~40%, and the 1766T mutation in 24% of the HCC patients (Baptista et al., 1999). The

1753C1762T1764A1809T1812T combination occurred more often in HCC patients than the 1753C1762T1764A1766T1768A1809T1812T (20% vs 4%) (Baptista et al., 1999). Eighty percent of the sequences also contained the 1809T and 1812T variants in the Kozak region of the pre-C gene (Baptista et al., 1999). *In vitro* studies investigated the effects of the 1753C1762T1764A1766T mutation (Buckwold et al., 1996; Guarnieri et al., 2006) and the 1809T1812T mutation (Ahn et al., 2003). The former set of mutations enhanced viral replication and reduced HBeAg expression, and the latter caused a reduction in HBeAg expression. Together these mutations lead to even lower HBeAg levels, suggesting that individuals infected with a combination of the pre-C and BCP variants may experience earlier anti-HBeAg seroconversion (Ahn et al., 2003). It is necessary to correlate these variants with possible changes occurring elsewhere in the genome. This can be achieved by the analysis of complete genomes generated through whole genome amplification.

The BCP is also prone to more pronounced variation. Deletions in the CP are concentrated in the BCP between nts 1745-1780 (Gunther et al., 1996b; Kramvis and Kew, 1999; Preikschat et al., 2002; Du et al., 2007). Insertions in the BCP mainly occur between nucleotides 1760 and 1780 (Kramvis and Kew, 1999) and duplication of the sequences between nts 1641-1666 and between nts 1720-1789 have been reported (Gunther et al., 1996b; Fujiwara et al., 2005). While the clinical significance of these alterations remains unclear, studies have explored the effects of these mutations on viral gene expression and replication *in vitro* (Gunther et al., 1996b).

1.10.2.2. PRECORE AND CORE GENE

The pre-C gene codes for the HBeAg, which is not required for viral replication. However, the precore gene does contain elements essential for viral replication such as the start site for pgRNA synthesis, DR1 and most of the ε signal. The integrity of this gene is therefore essential for viral replication. The role of this gene (nt 1814 to nt 1901) was examined in the sera and tissue of HBV-related HCC in southern Africa using subgenomic PCR and DNA sequencing (Kramvis et al., 1998). The majority of the isolates were serotype *adw* (genotype A). A significantly higher prevalence of the 1862G→T (Val→Phe) mutation was revealed in HBeAg negative patients compared to HBeAg positive HCC patients. The 1888G→A silent mutation occurred more frequently, in 53% of the sera and 100% of the non-tumorous and tumorous tissue samples. 1862G→T and 1888G→A occurred together in approximately 10% of the patients. These mutations occur within the ε structure, which is critical for viral replication. 1862G→T occurs within the six nt bulge of ε within the priming region, whereas 1888G→A occurs in upper stem. In addition to being a silent mutation at the amino acid level the 1888G→A mutation stabilises ε and does not appear to impair viral replication compared to the 1862G→T mutation, which does impair viral replication in certain genotypes (Guarnieri et al., 2006) and may reduce HBeAg levels (Hou et al., 2002; Chen et al., 2008)

Several studies now consistently show that these changes occur exclusively in genotype A1 (Kramvis and Kew, 2007b; Kramvis et al., 2008). Not only is infection with subgenotype A1 associated with low serum levels of HBV DNA and low prevalence of HBeAg (Tanaka et al., 2004), but it is also associated with an increased risk of developing HCC compared to non-A genotypes (Kew et al.,

2005). These mutations may contribute to the hepatocarcinogenic nature of subgenotype A1. The analysis of the pre-C gene in HCC provides insight into the possible mechanism of HBV-induced pathology in Black Africans. These data have to be interpreted in context of changes occurring within the entire genome. Full genome analysis would confirm whether specific variations occurring in the pre-C gene occur with mutations occurring in the BCP and in the rest of the genome. Possible combinations of mutations may reveal virological characteristics, which may contribute to disease progression and could possibly be useful as predictive markers for hepatocarcinogenesis.

1896G→A results in TAG stop codon at residue 28 in the pre-C protein and prevents the translation of the HBeAg (Carman et al., 1989). 1896G→A also improves the stability of the ε signal in HBV strains which possess a T at position 1858. Genotypes B, D and E possess 1858T, while genotypes A and H possess 1858C, genotypes C and F can have either 1858T or 1858C. The 1896G→A mutation favours a T at position 1858 resulting in enhanced viral genome stability, pgRNA encapsidation and initiation of DNA synthesis (Lok et al., 1994). This variant is less prevalent in chronic carriers from South Africa and North America and more common in Asia and the Mediterranean because of differences in the prevalence of the different groups with or without 1858T.

1.10.2.3. X GENE

Alterations in CP/EH II affect the X ORF because of the overlapping nature of the HBV genome. The 1762T1764A mutation is prevalent in HCC in black Africans (Baptista et al., 1999) and in subgenotype A1 (Tanaka et al., 2004; Kramvis and Kew, 2007b). This mutation results in amino acid substitutions in the HBx protein

K130M and V131I, within the HBx transactivation domain (aa 61 to 142) (Takada and Koike, 1990a; Takada and Koike, 1990b; Wakita et al., 1991; Ni et al., 2000). Besides affecting viral replication and HBeAg expression these mutations have been implicated in hepatocarcinogenesis (Kao et al., 2003). The 1753T→C/A/G of which 1753T→C occurs most frequently, affects amino acid 127 within HBx and may also affect BCP activity (Takahashi et al., 1998). It has been proposed that these mutations are selected during chronic infection through immune pressure directed towards HBx (Takahashi et al., 1999). Deletions in the core promoter may partially or completely remove amino acids essential for *in vitro* transactivation HBx activity (aa134-aa141) (Arii et al., 1992; Gunther et al., 1996b).

1.10.2.4. POLYMERASE REGION

The HBV polymerase gene encodes a multifunctional protein central to the life cycle of HBV. This RNA- and DNA-dependent DNA polymerase has RNase and RT activity and is also responsible for priming DNA synthesis through a tyrosine-nucleotide linkage and packaging of the pgRNA. Before the use of nucleoside analogues for antiviral treatment, natural polymerase variants were rarely reported and as expected these mutations yielded defective virus. Lamivudine (2',3-dideoxy-3'-thiacytidine) is an effective inhibitor of HBV polymerase (Dienstag et al., 1995) and is widely used in the treatment of chronic HBV infection. Long term lamivudine therapy results in the emergence polymerase gene mutations. Mutations appear to be concentrated in the YMDD motif (Tyr-551-Met-552-Asp-553-Asp-554), which is located in the palm sub-domain of the polymerase (Torresi, 2002). This motif is considered the major catalytic site overlaps the 'a' determinant of the S ORF (Torresi, 2002). YVDD (Met552Val) and YIDD (Met552Ile) are the most commonly reported mutations associated with lamivudine resistance (Allen et al., 1998). These are independent mutations and YIDD (Tyr-551-Ile-552-Asp-553-Asp-554) is not an intermediate in the selection of YVDD (Tyr-551-Val-552-Asp-553-Asp-554). The Leu528Met substitution is another common change, often associated with Met552Val (Papatheodoridis et al., 2002). Lamivudine resistant variants are functional and pathogenic, but are less efficient than wild type viruses that re-establish themselves after the cessation of therapy (Ono-Nita et al., 1999a; Ono-Nita et al., 1999b). Met552Ile and Met552Val correspond to Ile195Met and W196S substitutions in the S ORF respectively, and may affect the structural integrity of the HBsAg (Torresi, 2002).

1.10.2.5. THE PRE-S AND SURFACE REGION

Naturally occurring pre-S variants containing point mutations and deletions prevail during advanced liver disease, which is characterised by lower HBV DNA levels (Fan et al., 2001). These pre-S variants have been implicated in the development of HCC (Sugauchi et al., 2003a; Huy et al., 2003). The expression of pre-S mutant HBs proteins often results in the establishment of ground glass hepatocytes (GGHs), types I and II. Histological GGHs are found to cluster together, this can be attributed to clonal and integrated expansion (Fan et al., 2001). GGHs are considered the histological hallmark of HCC and develop through accumulation of surface proteins in the ER. GGH type I is associated with pre-S1 deletions and type II with pre-S2 deletions, with or without a pre-S2 initiation codon mutation. Pre-S deletion variants are associated with decreased synthesis of MHBs and SHBs (HBsAg), impaired virion morphogenesis and HBsAg secretion (Gerner et al., 2003), and these imbalances contribute to the accumulation of protein in the ER. This has also been shown to induce ER stress pathways and ER stress associated DNA damage in hepatoma cell lines and HCC (Hsieh et al., 2004a).

ER stress and DNA damage leads to the activation of cell cycle regulatory pathways: apoptosis and cell cycle arrest (Hsieh et al., 2004b; Hampton, 2000; arcon-Vargas and Ronai, 2002). The p53 tumour suppressor is a key regulator of these cellular responses and responds to DNA damage and cellular stress by the transcriptional activation of several downstream genes. DNA-damage in particular induces p53 activation of the cyclin-dependent kinase inhibitor $p21^{Waf-1}$, which leads to a G_1 arrest allowing sufficient time for DNA repair to occur before re-entering the cell cycle to complete DNA synthesis and cell division (Xiong et al., 1993; el-Deiry et al., 1994). ER stress induced by the accumulation of protein also

results in the activation of several signalling pathways, which include the unfolded protein response (UPR). The prolonged activation of UPR can result in programmed cell death, which is mediated through the activation of caspase 12. Caspase 12 in turn activates caspase 9 and caspase 9 activates the terminal caspase 3, which leads to the execution of apoptosis. Although it has been established that pre-S variants induce cellular stress, it is not clear whether pre-S variants have the ability to interfere with these downstream regulatory pathways and if so at which junctures. It is also unclear as to whether subtle differences in pre-S variants have different effects. To date the molecular characterization and biological significance of pre-S variants has not been explored in HBV-induced HCC in southern Africa.

The S gene encodes the major surface antigen also known as HBsAg. This protein has functional and structural roles in the life cycle of HBV. HBsAg is critical for virion assembly and infectivity. In addition, it contains the highly immunogenic ‘a’ determinant at residues 124-147 (Carman et al., 1990), which is situated within a major hydrophilic loop (residues 100-160) of this protein. The surface exposure of this epitope makes it a major target for the host’s humoral immune response. Several vaccine escape mutants have been described. The residue 145Gly→Arg of HBsAg is the most prevalent mutation, while 126Thr→Asn, 129Gln→His, 133Met→Leu and 144Asp→Ala are also reported frequently (figure 1.10). These mutations are a major concern as infection of vaccinated individuals with vaccine escape mutants would result in acute infections. Insertions in HBsAg have been reported in Asians and occur between cysteines residues 121 and 124 (Yamamoto et al., 1994; Hou et al., 1995). Patients infected with vaccine escape mutants are

often positive for HBsAg and anti-HBs. These insertions may affect the fidelity of HBsAg commercial assays and result in false negatives, presenting a serious diagnostic dilemma (Yamamoto et al., 1994; Carman et al., 1995; Hou et al., 1995).

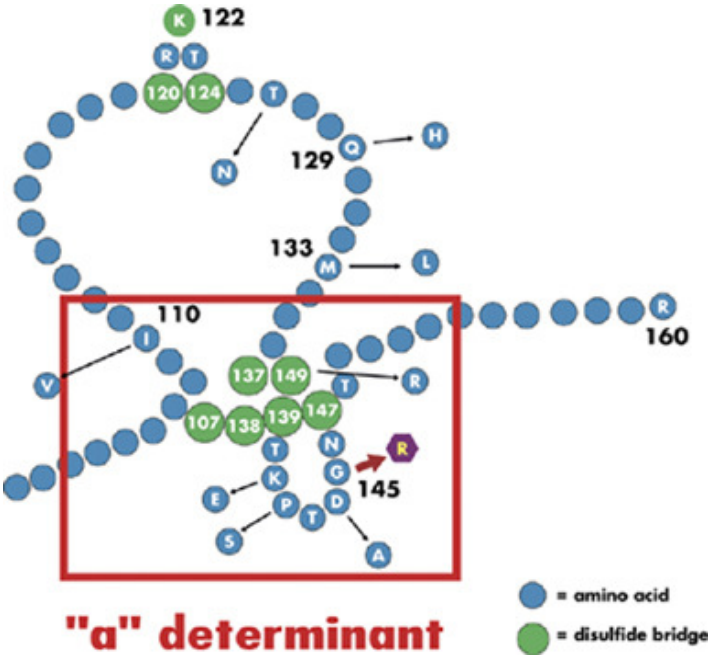


Figure1.10. The 'a' determinant of the HBsAg occurring within the major hydrophilic loop. The most prevalent vaccine escape mutant is indicated in purple at residue 145, the other reported vaccine escape mutants are indicated.

1.11. PRE-S FAMILY OF TRANSACTIVATORS

The family of pre-S regulatory proteins consists of 3' truncated MHBs ($MHBs^t$) and LHBs. The initial characterization of C-terminally truncated pre-S/S genes was based on the analysis of integrated HBV DNA in the human hepatoma cell line Huh 4 (Kekule et al., 1990) and HBV-induced HCC (Nagaya et al., 1987; Caselmann et al., 1990). The S ORF occurs within the single-stranded gap region of the HBV genome (Ganem and Varmus, 1987; Shih et al., 1987) and therefore may be preferentially targeted for integration (Nagaya et al., 1987; Caselmann et al., 1990; Schluter et al., 1994).

Several factors are required for transcriptional activator function of pre-S/S proteins. Their cytosolic topology, retention in the ER secretory pathway or expression in cytosol is essential (Hildt et al., 1993; Hildt et al., 1996a). The S region of MHBs has at least three (I, II and III) hydrophobic domains: aa 62-78, aa 135-153 and aa 224-281 (figure 1.4). The minimum prerequisite for transcriptional activator function of MHBs requires the loss of aa 224-281 within the C-terminal hydrophobic region (Lauer et al., 1992).

It is clear that integrated viral genes often encode truncated pre-S proteins (Schluter et al., 1994) and it is also possible for similar variants to be expressed from episomal HBV DNA, by the introduction nonsense mutations in the 3'-terminus of the S-gene (Hildt et al., 1995). These transactivators may induce hepatocarcinogenesis by constitutively activating tumour promoting pathways such as the protein kinase C/c-Raf-1/mitogen activated protein kinase/ extracellular signal-regulated kinase (PKC/c-Raf-1/MAPK/ERK) pathway that leads to the

downstream activation of gene transcription. Pre-S transactivators have not been investigated and identified in HBV-induced HCC patients in southern Africa.

1.12. SPLICE VARIANTS AND VARIANTS WITH INTERNAL POLY (dA) TAILS

HBV splice variants represent a subset of HBV genomes generated by the splicing of the pregenomic RNA at conserved splice donor and acceptor sites (Terre et al., 1991; Gunther et al., 1997a). HBV splice variants may play a role in viral persistence and HBV-induced disease. Splice genomes are more prevalent in the sera of acute patients that progressed to chronic infection or in those that already had chronic hepatitis (Rosmorduc et al., 1995). In addition, splice genomes preferentially accumulate within cells (Sommer et al., 2000) and can provide a reservoir for viral protein expression, leading to the intracellular accumulation of HBcAg and increased secretion of HBeAg (Rosmorduc et al., 1995). These characteristics may contribute to disease and viral persistence.

The prevalence of splice variants in HCC is not known. The majority of the studies, which correlate HBV heterogeneity with disease progression have used amplification of sub-genomic regions critical to viral replication and HBeAg expression (pre-C and BCP). As a result of this certain HBV variants are perhaps not detected and are therefore underrepresented in the literature.

In addition, certain variants can only be detected and described in their entirety by means of complete genome PCR amplification. Using techniques available to us it is not possible to directly clone the complete genome of HBV isolated from sera with low viral titres therefore it is necessary to first amplify the complete genome before cloning. This allows the characterization of the complete genomes of both major and minor quasispecies populations.

Another group of rare genomes containing internal poly (dA) tails have been detected using complete genome amplification. These variants probably emerge through unconventional replication steps and these have been found in patients with a range of clinical conditions including immunosuppressed chronic HBV carriers, fulminant hepatitis and an individual undergoing IFN-α therapy (Sommer et al., 1997). The prevalence of this variant in HCC is not known.

1.13. HBV DNA INTEGRATION

HBV DNA integration is not essential for HBV replication but is a hallmark of HBV infection. Necroinflammatory hepatic disease, which often accompanies HCC may promote HBV DNA integration and the accumulation of mutations in cellular genes. Integration also causes chromosomal instability in the vicinity of the integration site, and may lead to the aberrant expression of cellular and viral genes. Viral proteins expressed from truncated and rearranged inserted HBV sequences have been characterised as transactivators. These include the HBx protein and the family of pre-S2 containing HBs proteins. These viral genes are often conserved in integrated HBV DNA and their expression may lead to positive selection by clonal expansion of transformed cells (Schluter et al., 1994).

Chromosomally integrated HBV DNA is a common feature HBV-related HCC and has been reported in 90% of HCCs in Black Africans (Shafritz et al., 1981). HBV DNA integration appears to occur at random sites within the genome. It is not clear to what degree key cell cycle regulatory genes would be affected.

1.14. THE HBx TRANSACTIVATOR

HBx has been described as a promiscuous transactivator of cellular promoters (Yen, 1996). HBx is mainly located in the cytoplasm and does not bind DNA

directly but is able to affect the activation of several transcription factors, AP1, NFκB, SP1 and oct-1 by regulating cell signalling pathways that include the MAPK (Luber et al., 1993; Kekule et al., 1993; Benn and Schneider, 1994; Benn et al., 1996) and the Janus family of tyrosine kinase (JAK)/signal transducer and activators of transcription (STAT) pathways (Lee and Yun, 1998). The potential oncogenicity of HBx is reinforced by the observation that 90% of HBx transgenic mice develop HCC (Kim et al., 1991). In addition, the majority of HBV-related HCCs contain integrated X transactivator sequences (Schluter et al., 1994).

1.15. RECOMBINANTS

Exchange between HBV sequences by recombination allows for evolutionary leaps vastly exceeding the variation achieved through natural nucleotide drift. In HCC, homologous recombination was shown in episomal and integrated HBV DNA (Georgi-Geisberger et al., 1992). Sequential analysis of emerging quasispecies during chronic HBV infection demonstrated recombination between wild-type and mutant viruses (Tran et al., 1991). Furthermore recombination has been demonstrated between genotypes. The co-circulation of HBV genotypes occurs in various geographical regions and presents the opportunity for infection with multiple genotypes. Co-infection in turn facilitates the exchange of genetic material between HBV strains within an individual. Genotype B and C recombinants subgenotype B2 (Ba) are found throughout Asia, whereas genotype B without recombination, subgenotype B1 (Bj) is found in Japan (Sugauchi et al., 2002). Subgenotype B2 is associated with a higher incidence of HCC and younger age in Taiwan, whereas subgenotype B1 was associated with lower incidence of HCC and older age in Japan (Orito et al., 2005). The predominant subgenotype in Tibet is a C/D recombinant (Cui et al., 2002). In southern Africa A/D recombinants

have been reported in asymptomatic chronic carriers and this is not entirely surprising, because genotypes A and D co-circulate in this region (Owiredu et al., 2001a). The latter analysis used overlapping subgenomic PCR analysis to reconstruct complete genome sequences. It would be useful and interesting to determine whether recombinants occur in HCC patients from this geographical region, and this could readily be achieved using complete genome amplification and computational analysis.

1.16. HBV SEQUENCE VARIATION IN SOUTHERN AFRICA

Southern Africa has one of the highest incidences of HBV-related hepatocellular carcinoma in the world (Parkin et al., 1999). Three original South African studies have investigated the relationship between HBV and HCC in Black Africans. In a case control format, it was established that HCC patients were predominantly infected with subgenotype A1 and more significantly, individuals infected with subgenotype A1 were 4.5 times more likely to develop the tumour compared to those infected with non-A genotypes. Earlier studies investigated the prevalence of mutations in the pre-C gene and BCP. When predominantly *adw* (genotype A) isolates from HCC patients were analysed, Kramvis et al found a correlation between the 1862G→T and HBeAg negativity (Kramvis et al., 1998). This mutation occurs within the bulge of ε and can reduce viral replication (Guarnieri et al., 2006) and HBeAg expression (Hou et al., 2002; Chen et al., 2008). The 1888G→A mutation, which stabilizes the pgRNA by converting a wobble base U:G into a U:T within the upper stem of ε signal was also prevalent in the South African HCC isolates (Kramvis et al., 1998). A follow up study of the BCP, showed that double mutation 1762A→T1764G→A was more prevalent in patients with HCC (66%) than asymptomatic chronic carriers (11%) (Baptista et al., 1999). In addition, this

was the first study to reveal the presence of the 1809G→T1812C→T variants in southern African isolates. It has since been established that the 1809G→T, 1810C→T, 1812C→T, 1862G→T and 1888G→A variants are unique to subgenotype A1 (Sugauchi et al., 2004a; Kimbi et al., 2004; Kramvis and Kew, 2007b; Kramvis et al., 2008) and may account for the distinctive virological characteristics of this genotype (Kimbi et al., 2004; Kramvis and Kew, 2007a). The characteristically low serum levels of HBV DNA and the low prevalence of HBeAg of subgenotype A1 may contribute to viral persistence and severity of HBV-induced disease, respectively.

To date studies conducted in southern African Black patients used subgenomic PCR to reveal important findings about HBV genotype and specific mutations within the regulatory regions of HBV in relation to end stage liver disease. The utilization of subgenomic PCR meant that mutations in one part of the genome could not be correlated to mutations in another region. To gain a clearer understanding of HBV-induced disease, genotypic variations and mutations, specifically related to HCC, have to be viewed in their entirety on a single genome. This can be achieved by complete genome amplification of HBV isolated from the serum of HCC patients, as these sequences are representative of the replicating virus pool within the patient.

Despite the high prevalence of HBV in Africa and the huge associated HBV-induced disease burden, there are only 84 complete HBV genomes of African origin that have been deposited into international databases: 45 genotype A isolates, 10 genotype D isolates and 29 genotype E isolates (Kramvis and Kew, 2007a). These have mainly been isolated from blood donors, chronic carriers and

patients with acute hepatitis or fulminant hepatitis. Complete genome isolates of HBV from HCC patients are not available. To gain a clearer understanding of the relationship of genetic variants and mutations with HCC, it is advantageous to study the sequence of the full HBV genome.

The aim of the present study was to amplify the complete genome of HBV isolated from South African patients with HCC in order to characterise the genome in its entirety.

1.17. AIMS AND ORGANIZATION OF THE STUDY

The aims of this study were to:

- Extract HBV DNA from sera of hepatocellular carcinoma patients.
- Amplify and screen full genome isolates from hepatocellular carcinoma patients.
- Characterise the genomes at a molecular level.
- Identify changes in the open reading frames.
- Determine the heterogeneity of mutants occurring within and between patients.
- Link the molecular changes to genotype and disease.
- Determine whether molecular changes are related to genotype and/or disease manifestation

The results of this study are presented in Chapter 3 within five sections:

1. **SEQUENCE CHARACTERISTICS AND PHYLOGENETIC ANALYSIS OF HBV FROM BLACK SOUTHERN AFRICAN HEPATOCELLULAR CARCINOMA PATIENTS**.

   This section involved the amplification and molecular characterization of complete HBV isolates from patients with HBV-induced HCC. Phylogenetic analysis was used to determine the genotype/subgenotype. The isolates identified in the patients with HCC were compared to subgenotype A1 GenBank sequences isolated from asymptomatic chronic carriers and patients with acute hepatitis. This allowed a comparison of the amino acids in the four ORFs and nucleotides changes in relevant *cis*-regulatory regions.

2. **SIMULTANEOUS ALTERATIONS IN THE BASIC CORE PROMOTER, CORE GENE AND SURFACE GENE.**

   Five complete genomes were sequenced from a single patient. The molecular characterization of these five clones allowed for the identification of differences and similarities within the four ORFs within a single patient. One of the clones had a complex set of mutations within the basic core promoter, the core gene and the surface gene.

3. **IDENTIFICATION OF RARE VARIANTS IN HEPATOCELLAR CARCINOMA.**

   The isolates from 2 patients displaying rare variants were fully characterised. In the first patient four complete genomes were sequenced: two genomes had a wild-type length, a third has a pre-S2 gene deletion and the fourth was a rare variant with an internal poly (dA) tail. The complete genomes of 3 clones were characterised from a second patient: one genome had a wild-type length, the second had a pre-S2 deletion and the third was a splice variant.

4. **ANALYSIS OF THE SURFACE OPEN READING FRAME.**

   This section focused on the alterations identified in the pre-S and S regions. In addition to the complete surface gene sequences obtained from complete genomes; subgenomic amplification was carried out on the complete surface gene in additional patients. These sequences were translated, aligned and compared. Alterations were identified in the pre-S and S regions.

5. THE BIOLOGICAL EFFECTS OF PRE-S AND S GENE MUTATIONS.

Analysis of the surface open reading frame led to the identification of several mutations. These included pre-S internal deletions, pre-S1/pre-S2 initiation codon mutations and S gene nonsense mutations, which resulted in carboxyl end truncation. A number of mutants were chosen and cloned into an expression vector. These were transfected into hepatoma cells to determine the effect on anti-proliferative pathways. It must be emphasized that these data are preliminary and intended as a pilot screen of the clones generated.

CHAPTER 2

2.0 MATERIALS AND METHODS

2.1. SUBJECTS AND SERUM SAMPLES

Hepatitis B surface antigen (HBsAg)-positive sera were collected from 40 southern African Blacks with histologically proven hepatocellular carcinoma (HCC). These HCC patients were treated in a number of hospitals in Johannesburg. The sera were stored at -70℃ to prevent degradation of nucleic acids and protein. HBV serological markers were detected in the serum using commercially available kits (Abbott Laboratories, Chicargo, IL). The study was approved by the Human Ethics Committee of the University of the Witwatersrand (ethics number M080132 and recertification number M031033).

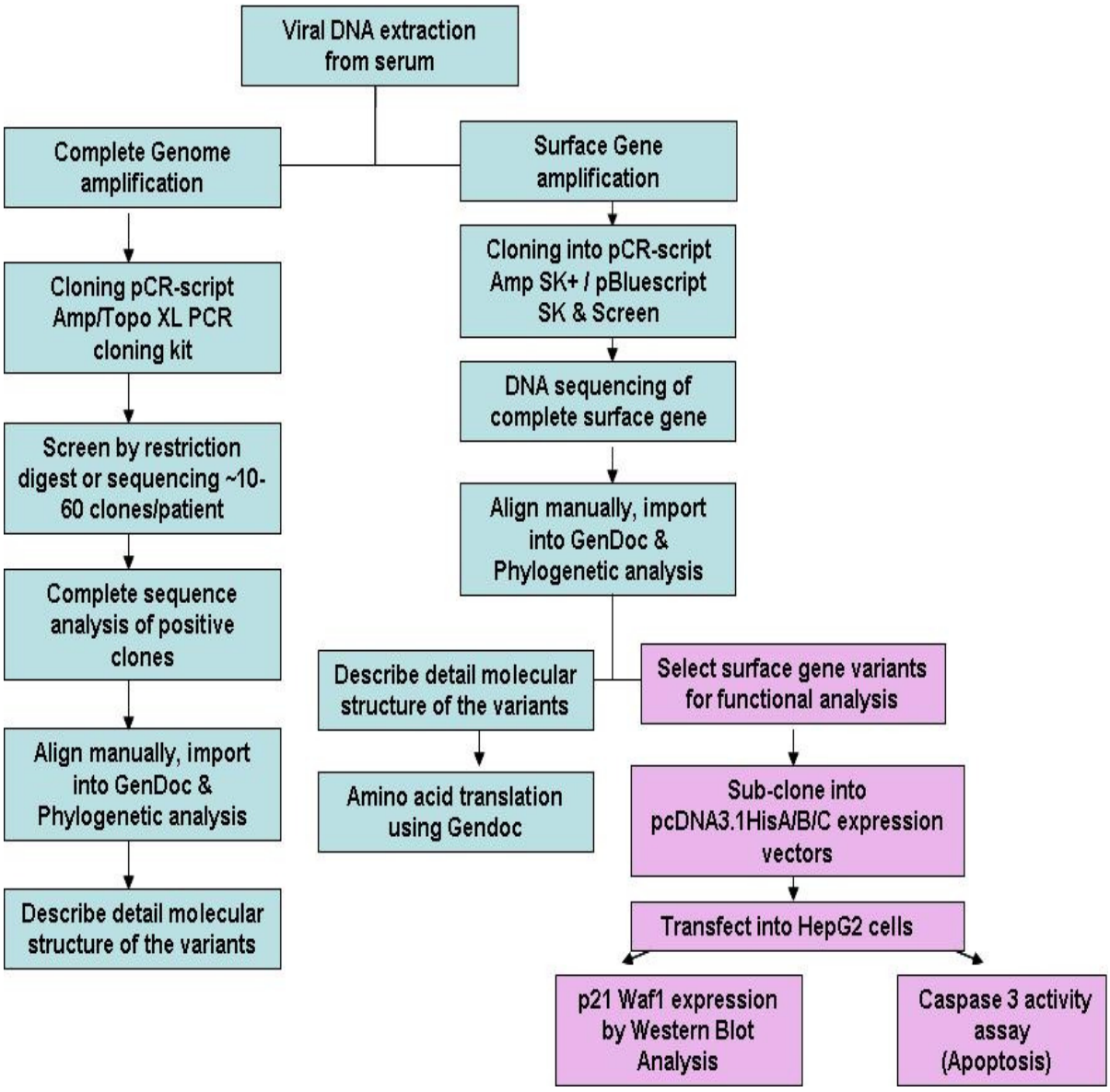


Figure 2.1. Flow diagram of methodology.

2.2. DNA EXTRACTION, AMPLIFICATION, CLONING AND SEQUENCING

2.2.1. DNA EXTRACTIONS

The QIAamp DNA kit (QIAGEN GmbH, Hilden, Germany) was used for nucleic acid purification (genomic, mitochondrial and viral) according to the manufacturer's instructions. A 100-200 µl aliquot of serum was used for the extraction. The DNA was eluted from the QIAamp spin column with 50 µl of 'best quality' water (BQW). Known HBsAg-positive sera, HBsAg-negative sera and BQW were used as controls for the extraction procedure.

2.2.2. DNA AMPLIFICATION

2.2.2.1. COMPLETE GENOME AMPLIFICATION

The Expand High Fidelity PCR system (Roche Diagnostics, Mannheim, Germany) was used for the amplification of complete HBV genomes. The reaction mixture consisted of 4.5 µl of 10X Expand High Fidelity (HF) buffer with 15 mM $MgCl_2$, 0.4 µl of 25 mM dNTP mix, 5 µl of 10 µM primer P1 and P2 or 1800 (-), and 5 µl of extracted DNA made up to 45 µl with distilled water. The enzyme mixture was made up of 0.5 µl HF buffer, 0.75 µl of Expand HF enzyme (3.5 units/ µl) and 3.75 µl of distilled water. The reaction was preheated to 95˚ C for 5 minutes and then cooled to 60˚C at which point 5 µl of enzyme mixture was added. This was followed by 40 cycles, cycling profile indicated in Table 2.1. Both HBV positive and negative controls were included, the latter consisting of BQW instead of DNA in the PCR reaction mixture. The reaction was carried out in a Perkin Elmer GenAmp PCR system 9600 / 2700 (Norwalk CT) or a Robocycler gradient 40 (Stratagen, La Jolla, CA).

2.2.2.2. SUB-GENOMIC AMPLIFICATION

Sub-genomic semi-nested PCR was performed on the complete surface gene. The reaction mixture for the first round consisted of 2.5 µl of 10X NH_4Cl reaction buffer, 1.5 µl of 50 mM $MgCl_2$ buffer, 0.2 µl HotStar Taq DNA Polymerase (5 units / µl) (QIAGEN, GmbH, Hilden, Germany), 0.2 µl of 25 mM dNTP mix, 2.5 µl each of 10 µM primer 2408 (+) and 1011 (-), 2.5 µl of DNA and made up to 25 µl with distilled water. For cycling profile see Table 2.1. The reaction mixture for the second round consisted of 5 µl of 10X KCl buffer containing 15 mM $MgCl_2$, , 0.4 µl of 25 mM dNTP mix, 0.2 µl HotStar Taq DNA Polymerase, 5 µl of each 10 µM primer 2800 (+) and 1011 (-), 5 µl of first round PCR product and made up to 50 µl with distilled water. For cycling profile see Table 2.1. Once positive samples were identified, the PCR was repeated with the Expand High Fidelity PCR system and the amplicons prepared for cloning.

2.2.2.3. SEMI-NESTED POLYMERASE CHAIN REACTION (PCR)

Subgenomic PCR was performed across the precore/X gene region. The reaction mixture consisted of 5 µl of 10X NH_4Cl reaction buffer, 3 µl of 50 mM $MgCl_2$ buffer, 0.2 µl HotStar Taq DNA Polymerase (5 units / µl) (QIAGEN, GmbH, Hilden, Germany), 0.4 µl of 25mM dNTP mix, 5 µl each of 10 µM primer 1661 (+) and 2005 (-), 2.5 µl of DNA and was made up to 50 µl with distilled water. For cycling profile see Table 2.1. At least 150 µl of amplified product was generated for the necessary samples and prepared for cloning and sequencing.

Table 2.1. PCR Primers and Conditions used for amplification

Primer	Amplification Type	Genome Position [b]	Sequence 5'-3'	Denaturation	Annealing	Extension	Size[c]	Reference
P1 P2	Complete Genome	1821-1841 1825-1806	*5'* TTTTTCACCTCTGCCTAATCA3' *5'*AAAAAGTTGCATGRTGMTGG3'	95˚C 40 sec	60˚C 90 sec	68˚C 180 sec [d]	3 221	Gunther *et al*., (1995)
P1 1800 (-)	Complete Genome	1821-1841 1800-1775	*5'*TTTTTCACCTCTGCCTAATCA3' *5'*AGACCAATTTATGCCTACAGCCTCCTA*3'*	95˚C 40 sec	60˚C 90 sec	68˚C 180 sec	3 200	Gunther *et al*., (1995)
2408 (+) 1011 (-)	Subgenomic	2408-2437 1011-979	5'TCTCAATCGCCGCGTCGCAGAAGATCTCAA3' 5'CAAAAGACCCACAATTCTTTGACATACTTTCCAAT3'	94˚ C 30 sec	58˚C 60 sec	72˚C 180 sec	1 800	-
2800 (+) 1011 (-)	Subgenomic	2800-2833 1011-979	5'CACGTAGCGCATCATTTTGCGGGTCACCATATTCT 3' 5'CAAAAGACCCACAATTCTTTGACATACTTTCCAAT3'	94˚ C 30 sec	59˚C 50 sec	72˚C 90 sec	1 200	-
1661 (+) 2005 (-)	Subgenomic	1661-1678 2007-1991	5'GACTCTTGGACTCTCAGC3' 5'GCTGAGGCGGTATCTAGAAG3'	94˚ C 30 sec	52˚C 60 sec	72˚C 60 sec	350	-
HBV *Bam* Hi HBV Not i	Subgenomic	2812-2829 1006-985	5' *ATA GGA TCC* CATT TT GCG GGT CAC CAT 3' 5' *CTC GAG CGG CCG* CGA CCC ACA ATT CTT TGA CAT AC 3	94˚ C 30 sec	52˚ C 30 sec	72˚ C 30 sec	1 200	-

Abbreviations (+), sense; (-) antisense

[a] All the amplifications began with an initial denaturation at 94 ˚ C / 95 ˚ C for 2 minutes and ended with an extension at 72 ˚ C for 10 minutes

[b] Denotes the nt position of hepatitis B virus *adw* genome (GenBank accession # AY233276) where the *Eco*RI cleavage site is position 1

[c] Size of amplicon in base pairs (genotype A)

[d] Extension time increased by 120 sec every 10 cycles

Nucleotides in *italics* are restriction enzyme sites

R=A/G

M=G/A/T/C

2.2.3. DETECTION OF AMPLICONS

A 10 µl aliquot of the amplicon was analysed by agarose gel electrophoresis. The percentage of the agarose used (0.8-2 %) was determined by the expected amplicon size. Commercially available DNA size markers, 100 bp or 1 kb ladder (Promega Madison, USA) or Lambda DNA digested with *Pvu*I or *Hind*III were used as DNA markers. PCR positive and negative controls were always included on the gels. HBV DNA from HBsAg positive patients known to be HBV PCR positive were used as positive controls. BQW or DNA extracted from HBsAg negative patients were used as negative controls. The gels were electrophoresed in 1X TBE (Appendix A) at 40 mA for 30-60 minutes. Ethidium bromide was included in the preparation of the gel at a final concentration of 1 µg/ml. This allowed for viewing of DNA under UV illumination. The PCR products were stored at -20°C until further analysis.

2.2.4. CLONING OF AMPLICONS

In order, to ensure that a single HBV genome could be sequenced and that there was sufficient template available, amplicons were cloned into pCR-Script Amp cloning kit (Stratagene, La Jolla, CA) or the TOPO XL PCR cloning kit (InVitrogen Life Technologies, USA). For pCR-Script Amp cloning purposes the amplicons were column purified and blunt-end ligated into the pPCR-Script Amp SK+, recombinants were transformed into XL1-blue super competent bacterial cells according to the manufacturer's instructions. The transformed cells were plated onto ampicillin enriched agar plates containing 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) and isopropyl-beta-D-thiogalactopyranoside (IPTG) (Appendix A) for colour selection of positive clones. Cloning into the pCR-XL-TOPO vector required gel purification of the amplicons. The amplicons were electrophoresed on

crystal violet stained agarose gels. The gel purification components were provided in the kit. The recombinants were transformed into TOP 10 supercompetent bacterial cells according to the manufacturer's instructions. The transformed cells were plated onto kanamycin enriched agar plates.

For amplification of plasmids containing HBV DNA inserts, positive colonies were inoculated into 2-5 ml of Luria broth (LB) containing the appropriate antibiotic, either ampicillin or kanamycin (Appendix A) and incubated at 37˚C overnight shaking at 200 rpm. The plasmid DNA was then extracted using a QIAprep Miniprep kit protocol (QIAGEN GmbH, Hilden, Germany). Before the DNA extraction, glycerol stocks were prepared for all the bacterial cultures by adding 500 µl of culture to an equal volume of sterilized glycerol and these were stored at -70˚C.

To identify pPCR-Script Amp SK+ recombinants with HBV inserts, plasmid DNA was digested with *Pvu*II (MBI Fermentas GmbH, Germany). The digestion mixture consisted of 1 µl of *Pvu*II (10 units / µl) , 2 µl of Buffer G, 5 µl of plasmid and made up to 20 µl with distilled water. To identify pCR-XL-TOPO recombinants with HBV inserts, plasmid DNA was digested with *Eco*RV and *Sac* I (MBI Fermentas GmbH, Germany). The digestion mixture consisted of 0.5 µl of *Eco*RV (10 units / µl), 1 µl of *Sac* I (10 units / µl), 2 µl of Tango buffer (MBI Fermentas GmbH, Germany), 5 µl of plasmid and made up to 20 µl with distilled water. The digestion was carried out at 37˚C for 1-2 hours, the products were electrophoresed in 1% agarose gels stained with ethidium bromide and visualised under UV illumination. Fragments of interest varied between 3.2 and 0.5 kb depending on the initial PCR product size and internal restriction sites. Plasmids selected for sequencing were grown in

larger volumes of LB and extracted using QIAGEN Plasmid Midi- or Maxi-kits (QIAGEN).

2.3. DNA SEQUENCING

2.3.1. MANUAL SEQUENCING

2.3.1.1. PREPARATION OF PCR TEMPLATES

5 - 10 µl of amplification product was pre-treated with two hydrolytic enzymes: exonuclease I (10.0 units/ µl) to remove sticky ends and primers, and shrimp alkaline phosphatase (2.0 units/µl) to dephosphorylate unincorporated nucleotides (Sequenase PCR product sequencing kit (United States Biochemicals, Cleveland, OH, USA)). Enzymatic pre-treatment was carried out at 37°C for 15 minutes and enzymes were heat-inactivated by incubation at 80°C for 15 minutes in a thermal cycler.

2.3.1.2. PREPARATION OF PLASMID DNA

Plasmid DNA was denatured using one-tenth the volume of a 2 M NaOH 2 mM EDTA solution at 37°C for 30 minutes. The DNA was then precipitated with one-tenth the volume of 3 M sodium acetate (NaAc) (pH 5.5) and 3 volumes of ice cold absolute alcohol. This DNA was recovered by centrifugation at 14 000 rpm in an Eppendorf centrifuge rotor for 15 minutes at 4°C. The pellet was washed with 70% ethanol to remove excess salt. The DNA pellet was allowed to air dry and resuspended in the appropriate volume of sterile water.

2.3.1.3. PRIMER ANNEALING AND LABELLING REACTION

3.2 pmol of primer was added to 200-500 ng of DNA. This mixture was denatured at 99°C for 2 minutes and then place in an ice/water bath for 5 minutes to facilitate annealing. The labelling reactions were prepared according to the manufacturers' instructions. A 10^{-1} dilution of labelling mix and 5 µCi α-^{35}S-dATP was prepared and added to the DNA. This was incubated at 37°C for 2 minutes. Termination was

carried out at 40℃ for 5 minutes. The sequences were analysed on an 8% polyacrylamide gel in glycerol tolerant buffer and autoradiographed using Kodak film (X-Omat, Sigma, St. Louis, MO, USA).

2.3.2. AUTOMATED SEQUENCING

Cloned amplicons were prepared for sequencing using the BigDye Terminator v3.0 Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, USA) according to the manufacturer's instructions and subjected to cycle sequencing on a GeneAmp PCR System 2700 thermal cycler (Applied Biosystems, Foster City, USA). The sequencing reactions were analyzed on an Applied Biosystems 377 automated sequencer (Applied Biosystems, Foster City, USA). All the sequences were analysed in both the forward and the reverse directions.

Table 2.2. Primers used for Sequencing

Primer	5' -3' Sequence	Genome position	Ta	Reference
192+	TCGTGTTACAGGCGGGGTTT	192-211	55	Takahashi *et al*., 1998
455+	CAAGGTATGTTGCCCGTTTG	455-474	55	Takahashi *et al*., 1998
684+	TGCCATTTGTTCAGTGGTTCGTAGGGC	684-710	60	Owiredu *et al*., 2001
1156+	CGGCAACGGTCAGGTCTGT	1156-1174	55	Takahashi *et al*., 1998
1763+	GGTCTTTGTACTAGGAGGCTG	1763-1783	58	Owiredu *et al*., 2001
2044+	CTCACCTCACCATACAGCACTCA	2044-2066	50	-
2335+	GCTCGGTACCAACTATACTGTTGTTAGACGA	2335-2353	55	Owiredu *et al*., 2001
2809+	CATCATTTTGTGGGTCACCAT	2809-2829	50	Owiredu *et al*., 2001
270-	AGAGAAGTCCACCACGAGTCTAGA	270-247	55	Owiredu *et al*., 2001
455-	CAAACGGGCAACATACCTTG	475-455	55	-
668-	GGCACTAGTAAACTGAGCCA	668-687	55	Takahashi *et al*., 1998
1573-	CCGGCAGATGAGAAGGCACAGACGG	1573-1549	60	Owiredu *et al*., 2001
1966-	GTCAGAAGGCAAAAACGAGAG	1966-1946	58	Owiredu *et al*., 2001
2498-	AAGCCCAGTAAAGTTTCCCACC	2498-2477	55	Owiredu *et al*., 2001
3094-	CCTGAGCCTGAGGGCTCCAC	3075-3094	55	Takahashi *et al*., 1998
T7	ATTAACCCTCACTAAAGGGA	772-791 on the pPCR-Script Amp vector	50	PCR-Script Amp instruction manual
T3	TAATACGACTCACTATAGGG	772-791 on the pPCR-Script Amp vector	50	PCR-Script Amp instruction manual
M13 +	GTAAAACGACGGCCAG	434-447 on the pCR-XL-TOPO vector	51	TOPO XL PCR cloning kit instruction manual
M13 -	CAGGAAACAGCTATGAC	204-220 on the pCR-XL-TOPO vector	50	TOPO XL PCR cloning kit instruction manual

3.2 pmol of primer was used sequencing per reaction.

Abbreviation (+) Sense; (-) Antisense

2.4. PHYLOGENETIC ANALYSES

Complete HBV genome sequences, complete surface gene and X-gene sequences were compared with corresponding sequences of HBV from GenBank. The alignments were edited manually in GeneDoc (Nicholas and Nicholas, 1997) or BioEdit (Hall, 2007). Phylogenetic trees were constructed with MEGA 4.0.1 (Tamura et al., 2007) using the neighbour-joining method, reliability was tested by bootstrap resampling with a minimum of 1 000 replicates. These data were also confirmed using PHYLIP (phylogeny inference package) version 3.5c (Felsenstein, 1995). DNAML (DNA maximum likelihood) alone and DNADIST consecutively with NEIGHBOR (neighbour-joining) were used to generate dendograms. SEQBOOT, DNADIST and NEIGHBOR were used for bootstrapping of 1000 datasets. CONSENSE was used to compute a consensus tree. Trees were visualized using Treeview Win 32 software program (Page, 1996). MEGA 4.0.1 (Tamura et al., 2007) and Dna*SP* 4.50.3 (Rozas and Rozas, 1995) were used to compute the pair-wise DNA distances and GraphPad Prism version 5 (Graves, 2003) was used for the statistical analysis.

2.5. CONSTRUCTION OF SURFACE GENE EXPRESSION VECTORS

PREPARATION OF THE PCDNA 3.1/HIS A, B AND C VECTORS

The pcDNA 3.1/His A, B and C (InVitrogen Life Technologies) are derived from pcDNA 3.1 and are 5.5 kb vectors designed for the high-level expression and purification of recombinant proteins in mammalian hosts. The vectors were supplied in three reading frames allowing for in frame cloning with a polyhistidine metal-binding tag. High-level gene expression is driven from the human cytomegalovirus immediate-early (CMV) promoter and an SV40 poly A signal prevents spurious expression. 10 µg of pcDNA 3.1/His A, B and C vectors were

subjected to double digests with *Bam* HI (100 units) and *Not* I (20 units) in 1X Buffer O (MBI Fermentas GmbH, Germany) or *Not* I (20 units) and *Xho* I (100 units) in 1X Buffer O (MBI Fermentas GmbH, Germany) depending on the cloning site requirements. Digests were routinely done in a final volume of 150 µl. All digests were carried out according to the manufacturer's instructions with modifications where necessary. The linearized vector was gel purified using the SV gel and PCR clean-up system (Promega Madison, USA). The eluted DNA concentration was determined by measuring the absorbance at 260 nm.

2.6. PREPARATION OF THE PRE-S/S GENE FRAGMENTS

Naturally occurring pre-S/S gene variants were amplified and cloned into pPCR-Script Amp SK+ or pCR-XL-TOPO (for sequencing purposes). These recombinants (Table 2.4) were then used for sub-cloning of the pre S gene variants into the pcDNA 3.1/His A, B and C expression vectors. 10µg of the recombinant vectors were subjected to partial double digests using *Bam*HI (5 units) and *Not*I (50 units) in 1X Buffer O (MBI Fermentas GmbH, Germany) in a final volume of 150 µl for 15 minutes at 37°C and heat inactivated at 80°C for 10 minutes. Where required 5 µg of recombinant vector was digested with *Hind*III (25 units) and *Pst*I (25 units) in 1X Buffer B (Promega Madison, USA) in a final volume of 100 µl for 15 minutes at 37°C and heat inactivated at 80°C for 10 minutes . The DNA fragments of interest were gel purified using the Wizard SV gel and PCR clean-up system (Promega Madison, USA). The DNA was eluted from the column in 50 µl of BQW. The eluted DNA was concentrated by precipitation with one tenth the volume of NaAc (pH 5.5) and 2 volumes of ice cold absolute ethanol. This was incubated at -20°C for 30 minutes and the DNA was recovered by centrifugation in an Eppendorf centrifuge at 13 000 rpm for 10 minutes. The pellet was washed with

100 µl of 70% ethanol and recovered by centrifuging at 13 000 rpm for 5 minutes at 4℃.The DNA was allowed to air dry and was reconstituted in the required volume of BQW determined by the ligation reaction.

2.7. LIGATION REACTION

T4 DNA ligase (MBI Fermentas GmbH, Germany) was used with the components listed in Table 2.3. A 12-16 hour incubation at 15℃ was carried out.

Table 2.3: Ligation Reaction Components

Reagent	Volume (µl)
Vector 100ng / µl	1
T4 DNA ligase 5U / µl	1
10X Ligation Buffer	2
50% PEG	2
DNA in sterile water	14
Final volume	20

Ten µl of the ligation reaction was incubated with 100 µl of super-competent cells on ice and then incubated at 42˚C for 90 seconds. The mixture was then incubated on ice for 2 minutes before adding 500 µl of SOC medium and growing at 37˚C at 250 rpm for 45-60 minutes. 50 µl was plated onto a selection plate and incubated at 37˚C for 12-16 hours. Colonies were selected for mini-preparation of plasmid DNA isolation as described in section 2.16.

2.8. SUB-CLONING STRATEGY

Three different approaches were used to insert the pre-S variants in frame into pcDNA 3.1/His A, B or C.

Sequence analysis revealed that samples 1.2, 2.2 and 12.9 were in the correct orientation for insertion via the *Bam*HI - *Not*I sites of pcDNA 3.1/His vector (Table 2.4).

Insert 24 was removed from pCR-XL-TOPO by the *Hind* III- *Pst*I and religated with pBluescript II SK (+) (Stratagene, La Jolla, CA) at the *Hind*III-*Pst* I site. This recombinant was transformed into DH5 alpha bacterial cells. A clone with the correct size insert was selected and sequenced. pBluescript II SK (+)Δ 24 was restricted with *Not*I and *Xho*I. This insert was inserted into pcDNA 3.1/His A via the *Not*I-*Xho*I restriction sites.

Sample 9.5 required PCR amplification using primers HBV *Bam*HI and HBV *Not*I, containing the appropriate cloning sites. The pPCR-Script Amp SK+ Δ 9.5 recombinant was used as the template. The reaction mixture consisted of 4.5 µl 10X Expand HF buffer with 15 mM $MgCl_2$, 0.4 µl of 25 mM dNTP mix, 5 µl of 10 µM primer HBV *Bam*HI and HBV *Not*I, and 5 µl (50 ng) of plasmid DNA made up to 45 µl with distilled water. The enzyme mixture was made up of 0.5 µl HR buffer, 0.75 µl of Expand HF enzyme (3.5 units/ µl) and 3.75 µl of distilled water. The reaction was preheated to 94˚C for 5 minutes and then cooled to 52˚C at which point 5 µl of enzyme mixture was added and amplified using a 35 cycle profile indicated in Table 2.1. Both HBV positive and negative controls were included, the latter consisting of water not DNA in the PCR reaction mixture. The reaction was carried out in a Perkin Elmer GeneAmp PCR system 2700 (Norwalk CT). The amplicon was gel purified and cloned into pCR-XL-TOPO as described above. A positive clone containing the correct size insert was selected. A *Bam*HI *Not*I fragment was then prepared from pCR-XL-TOPO Δ 9.5 and cloned into pcDNA3.1/His C.

Table 2.4. Surface gene expression vectors

Sample Mutation	Recombinants	Target pcDNA 3.1/His Vector	Pre-S Expression Vectors
1.2	pPCR-Script Amp SK+ Δ 1.2	pcDNA 3.1/His A	pcHBsAg 1.2
2.2	pPCR-Script Amp SK+ Δ 2.2	pcDNA 3.1/His C	pcHBsAg 2.2
9.5	pPCR-Script Amp SK+ Δ 9.5	pcDNA 3.1/His C	pcHBsAg 9.5
12.9	pPCR-Script Amp SK+ Δ 12.9	pcDNA 3.1/His B	pcHBsAg 12.9
24	pBluescript II SK (+) Δ 24	pcDNA 3.1/His A	pcHBsAg 24

2.9. CELL CULTURE

2.9.1. CELL LINE

The HepG2 cell line was suitable for this study as it exhibits most of the phenotypic and biochemical characteristics of differentiated hepatocytes (Knowles et al., 1984). The normal expression levels of both the p53 and RB tumour suppressor genes suggests that these cell cycle regulators remain intact (Puisieux et al., 1993). This cell line also has a p53 polymorphism at codon 72, leading to alternative coding for arginine or proline.

2.9.2. SUBCULTURE AND STOCKS

The HepG2 cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% fetal calf serum (FCS), 100 units/ml penicillin and 100 µg/ml streptomycin and grown at 37°C in a humidified atmosphere of 5% CO_2. A glycerol stock of frozen cells was removed from liquid nitrogen and rapidly thawed in a 37° C water bath in the tissue culture facility. The cells were transferred into a 25 mm^2 flask containing 5 ml of medium. The cells were left for approximately one week until confluency was achieved. The conditioned medium was removed, cells were washed twice with 2 ml of 1X PBS and then incubated with 1 ml of trypsin / EDTA at 37° C for 1-5 minutes. Once the cells appeared rounded upon viewing through a light microscope 3 ml of medium was added to the cells to neutralize the

trypsin / EDTA. The cell suspension was collected and recovered by centrifugation at 1 000 rpm for 5 minutes at room temperature. The supernatant was removed and the cells were thoroughly resuspended in 2 ml of medium. The cells were transferred to a 75 mm^2 flask. The HepG2 were typically seeded at a ratio 2:1 when the cells reached an 80% confluency. Cells were frozen down in DMEM supplemented with 10 % FCS and 20% tissue culture grade glycerol.

2.9.3. MYCOPLASMA TEST

The HepG2 cells were checked for infection with *Mycoplasma*. The HepG2 were grown in antibiotic free culture media for at least one week, these were trypsinized and plated onto sterile cover slips. The cells were allowed to attach to the cover slip overnight, 5 ml of fixing solution (Appendix A) was added to the media and left for a few seconds, followed by a gentle rinse with water. The water was poured off and the dish inverted to facilitate drying. While drying a drop of mounting fluid was placed onto a slide. Once dried the Hoeschst stain (Appendix A) was added to the cover slip with the cells for a maximum of 30 seconds. The stain was removed by rinsing with water, the cover slip was then placed face down onto the slide. A drop of microscope oil was placed onto the slide and the cells were viewed under a fluorescence microscope. The cell nuclei stained bright green, while small fluorescent particles distributed throughout the cytoplasm were indicative of *Mycoplasma* contamination.

2.10. CALCIUM PHOSPHATE TRANSFECTION

The HepG2 cells were plated at a density of 900 000 cells per 60 mm dish the day before the transfection. Three hours prior to the transfection, the medium was removed and replaced with fresh medium. The DNA and 2X HBS solutions were prepared separately in sterile tubes. The DNA and water were mixed, 2 M $CaCl_2$ was added and mixed.

Table 2.5: Calcium Phosphate transfection components

	60 mm	100 mm
Endotoxin free plasmid DNA	12 µg	20 µg
2 M $CaCl_2$	37 µl	62 µl
Nuclease-free water	make upto 300 µl	make upto 500 µl
2X HBS	300 µl	500 µl

The tube containing the 2X HBS was vortexed within a tissue culture hood. The prepared DNA solution was gradually added to the 2X HBS in a drop wise fashion. This was allowed to incubate at room temperature for 30 minutes. The DNA solution was vortexed again just prior to adding dropwise to the cells. The plates were gently agitated by hand to distribute the precipitate evenly over the cells. The plates were returned to the 37℃ CO_2 incubator for 18 hours. The transfection medium was then removed and replaced with fresh medium. The cells were then returned to the incubator and harvested at the 24 hour time point.

2.11. DETECTION OF Β-GALACTOSIDASE

The *E. coli* lac Z gene encoding β-galactosidase (β-gal) was employed as a histochemical reporter gene. β-gal was detected by using a widely used chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactosidase (X-gal). Upon cleavage of the glycosidic linkage within X-gal, colourless indoxyl monomers are produced. Indoxyl moieties form dimers which are non-enzymatically oxidized forming the indigo precipitate.

HepG2 cells were transfected with β-gal expression vector (pCMV-lac Z) to determine transfection efficiency. 12 µg of the β-gal expression vector was used to transfect a 50-70% confluent 60 mm dish of cells using the calcium phosphate method described above. After the 24 hour time point the cells were washed 3 times with 1X PBS and fixed with 1 ml of 4% paraformaldehyde solution (Appendix A) at room temperature for 20 minutes. Another 3 PBS washes were followed by the addition of the 1 ml X-gal stain (Appendix A) and incubated at 37°C for 10 minutes or 12-16 hours. The cells were observed under light microscope, as soon as indigo precipitate appeared the PBS washes were repeated. The cells were observed under fluorescence microscope (Zeiss Axio Skop 40).

2.12. DOXORUBICIN TREATMENT

Doxorubicin (DOX) (Sigma), an anthracycline antibiotic, is a widely used anticancer drug. DOX is believed to act by inducing DNA damage through topoisomerase II inhibition and free radical generation. The free radical generation induces oxidative DNA damage and apoptosis (Rebbaa et al., 2003). DOX was used to induce apoptosis in HepG2 cells in this study. Untransfected cells were treated with 2 µM DOX in the transfection experiments used for caspase 3 assay.

2.13. PROTEIN PURIFICATION

2.13.1. CELL LYSATE PREPARATION

The cells were washed three times with 2 ml of PBS (Appendix A). The media and PBS washes were collected in 12 ml tubes and centrifuged at 1 000 rpm for 5 minutes in a Super T12 using a SL-50T rotor (with adaptors) at 4℃ to pellet any cells in suspension, 200 µl of RIPA buffer containing 20µl of 10X Complete protease inhibitor (Roche Diagnostics, Mannheim Germany) was added to the dish while on ice. These adherent cells were collected using a rubber policeman to scrape them off the dish. The lysate and pelleted cells were transferred to a 1.5 ml eppendorf tube and incubated on ice for 30 minutes. The lysates were centrifuged at 12 000 rpm for 20 minutes at 4℃ then aliquoted and stored at -20℃ or -70℃ for long term purposes.

2.13.2. PROTEIN QUANTITATION

The protein concentration was determined using the BCA Protein Assay kit (Pierce, USA). This colorimetric assay was prepared in a clear 96 well plate. A maximum of 10 µl of protein was loaded per well in duplicate for samples and standards. A set of BSA standards, ranging from 125 µg/ml to 2000 µg/ml were used with each quantitation. 100 µl of detection solution (reagent A and B mixed at a ratio of 1:50) was added to each well. The plate was incubated at 37℃ for 1 hour. The absorbance was measured at 595 nm on an Anthos microplate reader 2001 (Anthos Labtec, Cambridge UK). Standard curves were drawn using Microsoft Excel and the protein concentrations were calculated from these curves.

2.14. WESTERN BLOT ANALYSIS

2.14.1. SODIUM DODECYL SULFATE POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE)

The percentage (12-15%) of acrylamide used was determined by the expected protein size. The Biorad minigel casting apparatus was used for pouring 1.5 mm thick gels, allowing for maximum sample volume loading. Protein loading dye was added to a 1X final concentration, further denaturation was achieved by a 10 minute incubation at 99℃ just prior to loading onto the gel. The gel was electrophoresed at 100 volts for the appropriate times (1-2 hours). Protein migration was monitored by including a peqGold prestained protein marker IV (PEQLAB Biotechnology) as a size maker.

2.14.2. TRANSFER

The SDS-PAGE gels were removed from the glass plates. The gel was sandwiched between sponges, filter paper and Hybond ECL nitrocellulose membrane (Amersham Pharmacia Biotech, Buckinghamshire, England) on the appropriate side of the transfer cassette. Once the transfer cassette was assembled it was placed into the transfer tank with transfer buffer and ice block for cooling purposes. The protein transfer was performed at 200 volts for 1 hour.

2.14.3. PONCEAU S (MEMBRANE) AND COOMASSIE BLUE (GEL) STAINS

After the transfer was completed the membrane was treated with Ponceau S stain for 10 minutes with gentle shaking. Once excess stain was removed with distilled water, protein loading and transfer efficiency could be accessed. The membranes were photocopied or scanned before being prepared for analysis. The gels were also recovered and stained with Coomassie stain for 10 minutes and placed in

destain for 12-16 hours. The gels were also recorded by scanning as drying 1.5 mm thick gels were not feasible.

2.14.4. PRIMARY ANTIBODY

After the Ponceau S staining the membrane was washed 3 X for 1 minute with 1X TBS 0.1 % Tween then placed in 1 X TBS 0.1 % Tween with 5 % non fat milk (blocking solution) for 1 hour at room temperature with shaking. The membrane was then washed 3 X for 10 minutes with 1X TBS 0.1 % Tween and then incubated in primary antibody solution prepared in blocking solution with the indicated dilutions (Table 2.6) for 12-16 hours with shaking at 4 ℃.

2.14.5. SECONDARY ANTIBODY

After the primary antibody solution was poured off the membrane was washed 3 X 10 minutes with 1 X TBS 0.1% Tween. The appropriate HRP-conjugated secondary antibody dilution (Table 2.7) was prepared in the required buffer (Table 2.7) and allowed to incubate for 1 hour at room temperature with shaking. This was followed by 3 X 10 min washes with 1 X TBS 0.1% Tween to remove excess antibody. The proteins were detected by chemiluminescence using SuperSignal West Dura Extended Duration Substrate (Pierce) according to the manufacturer's instructions. The membranes were exposed to medical Xray film (Agfa, USA) for 1-10 minutes and developed in Structurix G128 developer (Agfa) and fixed in Rapid fixer G333c (Agfa) according to the manufacturer's instructions.

2.14.6. MEMBRANE STRIPPING

When necessary the membranes were stripped by placing in stringent stripping buffer for 30 minutes at 50℃ with intermittent shaking. Three 10 min 1 X TBS 0.1 % Tween followed to ensure the removal of all the β-mercaptoethanol. A 1 hour block was carried out and the membrane was prepared as before for the analysis of other proteins.

Table 2.6: Antibodies used in Western Blot analysis

Name	Antibody		Recognition site	Dilution	Incubation Solution	Supplier
p21 (H-164): sc-756	Primary	rabbit polyclonal	amino acids 1-164	1:200	5% block	Santa Cruz Biotechnology, INC
p38	Primary	rabbit polyclonal	amino acids 343-360	1:5000	1X TBS 0.1% Tween	Sigma
Goat Anti-rabbit IgG (Htc)-HRP conjugate	Secondary		p21 and p38	1:5000	5% block	Bio-Rad Laboratories, Inc

2.15. CASPASE 3 ASSAY

Caspase 3 activity was used to measure apoptosis. When cell lysates containing caspase 3 were added to the fluorogenic substrate aspartate-glutamate-valine-aspartate-7-amino-4-methyl coumarin (DEVD-AMC) cleavage resulted in the release of the fluorogenic AMC.

HepG2 cells transfected with the pre-S expression vectors were harvested after 48 hours. The cells were washed 3 times with PBS and 200 µl of lysis buffer was added to a 60 mm dish. The PBS and conditioned medium were collected separately in 12 ml tubes and centrifuged at 1 000 rpm at room temperature in a Beckman TJ-6 centrifuge centrifuge check centrifuge and rotor (Beckman Coulter, UK). The pelleted cells were added to the lysates. The lysates were stored on ice. These were always prepared freshly. The assay was prepared in an opaque 96 well plate. 7 µl (100 µM) of the fluorogenic DEVD-AMC substrate followed by 43 ul of caspase reaction buffer (Appendix A) were added to each well. This was covered and placed at 37℃ for 5 minutes before adding 50 µl of cell lysate. The plate was then placed at 37℃ for 2-12 hours. The plate was analyzed using the Fluorescence Spectrophotometer (Cary Eclipse, Varian) with excitation 380 nm and emission at 445 nm.

2.16. PLASMID DNA EXTRACTIONS FROM BACTERIAL CELLS

Small, medium and large scale plasmid preparations were done using the QIAGEN Plasmid Mini, Midi and Maxi kits (QIAGEN Inc, Hilden, Germany) respectively. The EndoFree Plasmid Maxi kit was used for the preparation of endotoxin free plasmid DNA (QIAGEN Inc, Hilden, Germany). 10-100 ng of plasmid DNA was routinely transformed into DH5 α or XL1-Blue competent cells by heat shock. The transformed bacterial cells were then grown in the required volume of Luria broth (LB) (Appendix A) at 37℃ for 12-16 hours with vigorous shaking (~250 rpm). 150 ml of bacterial cultures were recovered at 2 000 rpm in a Beckman centrifuge using a JA-10 rotor for 15 minutes at 4℃. The plasmid DNA was then isolated according to the manufacturer's protocol. The DNA concentration was determined by measuring the absorbance at 260 nm. The

A_{260}:A_{280} ratio was used to determine DNA purity, a ratio equal to or above 1.8 was of sufficient purity and adequate for transfection into mammalian cells. A sample of the DNA (~1µg) was electrophoresed on a 0.8% agarose gel at 40 mA, in 1X TBE buffer for 30-60 minutes to assess the quality of the DNA.

2.17. PREPARATION OF COMPETENT BACTERIAL CELLS

The DH5 α or XL1-Blue bacterial strains were routinely prepared in the laboratory. A single colony from an LB plate was used to inoculate 5 ml of LB. The culture was grown at 37℃ for 12-16 hours with shaking at at 250 rpm. This was used to inoculate 200 ml of LB at a final dilution of 1:100. The cells were allowed to grow to a growth index of 0.2 - 0.4 (A_{600}). The culture was then placed in 50 ml Corex tubes and placed on ice. The bacterial cells were gently recovered at 2 000 rpm in a Sorvall Super T12 using a SL-50T rotor for 30 minutes at 4℃. The supernatant was discarded and 2.5 ml of transforming buffer A (Appendix A) (4℃) was added to each pellet. The cells were gently resuspended using a sterile 5 ml pipette. The suspension was incubated on ice for 20 minutes. The suspension was centrifuged at 900 rpm for 10 minutes at 4℃. The pellet was resuspended in 500 µl of transformation buffer A (Appendix A) and stored at -70℃ in 50 µl aliquots.

CHAPTER 3

3.0 RESULTS

3.1 SEQUENCE CHARACTERISTICS AND PHYLOGENETIC ANALYSIS OF HBV FROM BLACK SOUTHERN AFRICAN HEPATOCELLULAR CARCINOMA PATIENTS

3.1.1. AMPLIFICATION, CLONING AND SEQUENCING

Both full-length amplification (Gunther et al., 1995) and subgenomic PCR were used to amplify HBV DNA from DNA extracted from a total of 40 patients with HBV-induced HCC. Ninety percent of the patients (36/40) were positive for HBV PCR by the full-length and/or subgenomic PCR. Thirty-three (82.5%) of the 40 extracts were positive for HBV DNA using the full-length PCR. The amplicons ranged in size from 0.5 kb to 3.2 kb (Table 3.1.1). From 9 patients, twenty six complete genomes were successfully cloned (Table 3.1.1). The length of the clones ranged from 1 120 bp to 3 253 bp. Twenty eight of the DNA samples were amplified using surface gene subgenomic PCR and 23 (82.1%) yielded positive results (Table 3.1.1). The surface gene amplicons varied between 983 bp to 1 203 bp in size. Twenty one amplicons, derived from 10 patients, were successfully cloned. Patients 15, 27 and 30 were negative for full-length PCR but positive for the surface gene subgenomic PCR (Table 3.1.1).

Table 3.1.1. PCR and Cloning HBV Data of HCC Patients

			Full Genome				Surface Gene			
Patient No.	Sex	Age	PCR Product sizes (kb)	No. of clones analyzed	Clone	Size of Genome (bp)	PCR	No. of clones analyzed	Clone Gene	Size of Surface (bp)
1	M	28	2 - 0.7	0		-	+	1	1.2	1 170
2	M	48	3.2	0		-	+	4	2.2	1 161
									2.4	1 161
									2.7	1 155
									2.9	ID
3	M	40	3.2 - 1	0		-	+	3	3.1	1 203
									3.2	1 203
									3.3	1 203
4	M	NA	3.2- 1	3	4.1	3 203	+	0		1 185
					4.4	3 203				1 185
					4.6	3 170				1 153
5	M	29	3	0	-	-	-ve	0		-
6		40	3.2 - 0.5	4	6.1	1 120	+	0		-
					6.2	3 212				1 194
	F				6.5	3 221				1 203
					6.13	3 221				1 203
7	F	27	-ve	0	-	-	-ve	0		-
8	M	NA	0.7 & 0.5	0	-	-	-ve	0		-
9	M	NA	3	4	9.16	3 002	+	1	9.5	984
					9.32	3 002				984
					9.36	3 002				984
					9.43	3 002				984
										984
10	M	75	3.2 - 1	3	10.17	1 579	+	0		-
					10.24	3 179				1 161
					10.12	3 221				1 203
11	M	69	3.2	0	-	-	-ve	0		-
12	M	45	1.8 - 0.8	0	-	-	+	1	12.9	1 152
13	M	36	3.2 - 2	0	-	-	NI	0		-
14	F	37	3.2	0	-	-	NI	0		-
15	M	34	-ve	0	-	-	+	0		-
16	M	65	3.2 - 0.9	0	-	-	NI	0		-
17	M	42	-ve	0	-	-	NI	0		-
18	F	28	3.2	5	18.12	3 194	+	0		1 110
					18.1	3 252				1 203
					18.20	3 252				1 203
					18.31	3 252				1 203
					18.15	3 253				1 203
19	F	58	3.2	0	-	-	NI	0		-
20	F	51	-ve	0	-	-	NI	0		-
21	M	31	2	0	-	-	+	0		-
22	M	26	-ve	0	-	-	NI	0		-
23	M	24	3.2 - 0.7	0	-	-	+	1	23.1	1 203
24	F	42	3.2 -1	0	-	-	+	1	24.1	1 202
25	M	63	3.2 - 2	0	-	-	NI	0		-
26	M	50	3.2 - 1	1	26	3 098	NI	0		1 080
27	M	67	-ve	0	-	-	+	3	27.1	1 202
									27.2	1 203
									27.3	1 203
28	M	63	3 - 1.5	0	-	-	+	3	28.1	1 149
									28.2	ID
									28.3	1 149
29	M	38	1.9 - 1	0	-	-	+	3	29.2	1 202
									29.5	1 203
									29.6	1 203
30	F	60	-ve	0	-	-	+	0		-
31	M	46	3.2 - 2	0	-	-	+	0		-
32	M	55	3.2	0	-	-	+	0		-
33	M	20	3.2	0	-	-	NI	0		-
34	M	25	3.2 - 1	0	-	-	+	0		-
35	F	32	3.2	3	35.7	3 215		0		1 197
					35.4	3 215	NI			1 197
					35.12	3 215				1 197
36	M	47	3.2	1	36	3 167	NI	0		1 149
37	M	67	3.2	0	-	-	+	0		-
38	M	33	3.2	0	-	-	+	0		-
39	M	15	3.2 - 1	0	-	-	-ve	0		-
40	M	37	3.2	2	40.9	3 220	+	0		1 202
					40.18	3 221				1 203

Abbreviations: -ve, PCR negative result; +, PCR positive result
NA, data not available
NI, sample not included (full-length clone not available/serum depleted)
ID, Insufficient sequence data (background could not be eliminated)
kb, kilobases
bp, basepairs

3.1.2. PHYLOGENETIC ANALYSES

The complete genomes and the four ORFs individually of the 24 clones were aligned with 49 sequences from the databases and compared using phylogenetic analyses (figures 3.1 – 3.5). Phylogenetic analysis of genome 6.1 (1 120bp) and 10.17 (1 579bp) is presented separately in section 3.3 using the X ORF, as this was the only uninterrupted gene in these variants. Phylogenetic analyses of the subgenomic S gene clones are shown in section 3.4 .

All the HBV strains isolated from the HCC patients belonged to subgenotype A1. The strains clustered within subgenotype A1 when phylogenetic analysis of the four open reading frames was performed individually. Isolates from HCC patients did not separate phylogenetically from southern African asymptomatic chronic carriers, acute and fulminant hepatitis patients.

The serological subtype for all the HCC genotype A isolates was extrapolated from the sequences of the complete genome and of the surface gene. Eleven of eighteen patients (patients 1, 2, 3, 4, 6, 20, 12, 18, 24, 26 and 29) were subtype *adw2*; three (9, 28 and 36) were *ayw2* and three (23, 27 and 40) were *ayw*1. The serological subtype of clones derived from the same patient was always the same.

The mean pairwise DNA divergence was calculated using MEGA (intragroup) and Dna*SP* (intergroup). The statistical analysis was performed with GraphPad Prism. Mann-Whitney method was used to calculated *P* values. The mean DNA divergences (%) ± standard deviation and the range in parentheses (table 3.1.2). All the complete genomes of the HCC patients were included in these calculations except isolates from patients 26 and 35 because nucleotides 1800 -1821 were missing. Primer set P1 and 1800 (-) were used to amplify HBV DNA from these

two patients because amplification with P1/P2 was not successful. All the non-HCC subgenotype A1 isolates listed in figure 3.6 were used in the calculations.

The mean pairwise DNA divergence calculated for the subgenotype A1 strains isolated from the HCC patients was consistently higher than that of the subgenotype A1 isolates from asymptomatic chronic carriers, acute hepatitis patients and fulminant hepatitis patients and this was statistically significant (Table 3.1.2). This was evident for the complete genome sequences and the individual ORFs.

The pre-S1/S2 and pre-C/core genes of the strains derived from the HCC patients had the highest intragroup divergence compared to the other three genes. The intergroup divergence of the pre-S1/S2 was also the highest between the two patient groups (i.e. HCC vs. asymptomatic chronic carriers, acute hepatitis patients and fulminant hepatitis patients).

Table 3.1.2. Mean pair-wise DNA divergence (%) of complete genome sequences and individual ORFs of subgenotype A1 in different patient groups.

	Intragroup			Intergroup
	HCC	AS, AC & FH	HCC vs AS, AC &FH *P* value	HCC vs AS, AC & FH
Complete genome	3.18 ± 1.06 (0.00 - 4.70)	2.66 ± 0.88 (0.60-4.70)	<0.0001	3.06 ± 0.27 (0.46-5.54)
polymerase	2.96 ± 1.02 (0.00 - 4.40)	2.66 ± 0.82 (0.53-4.48)	<0.0001	2.91 ± 0.25 (0.41-6.02)
pre-S1/S2	4.02 ± 1.96 (0.00 - 7.70)	3.40 ± 1.25 (0.44-6.76)	0.0001	3.79 ± 0.31 (2.66-5.34)
HBsAg	2.17 ± 0.93 (0.00 - 4.30)	1.20 ± 0.48 (0.29-2.70)	<0.0001	1.80 ± 0.21(0.46-4.61)
pre-C/core	4.02 ± 1.39 (0.00-6.76)	2.44 ± 0.99 (0.31-5.09)	<0.0001	3.41 ± 0.33 (1.17-5.48)
X	1.94 ± 0.72 (0.00-3.60)	1.88 ± 1.39 (0.20 -6.80)	0.0025	2.08 ± 0.29 (1.05-3.08)

The subgenotype A1 sequences compared are those included in figure 3.6, 20 isolates from HCC patients excluding isolates from patients 26 and 35 and 18 isolates from the remaining groups including AF297623 and AF297625 and excluding M57663.

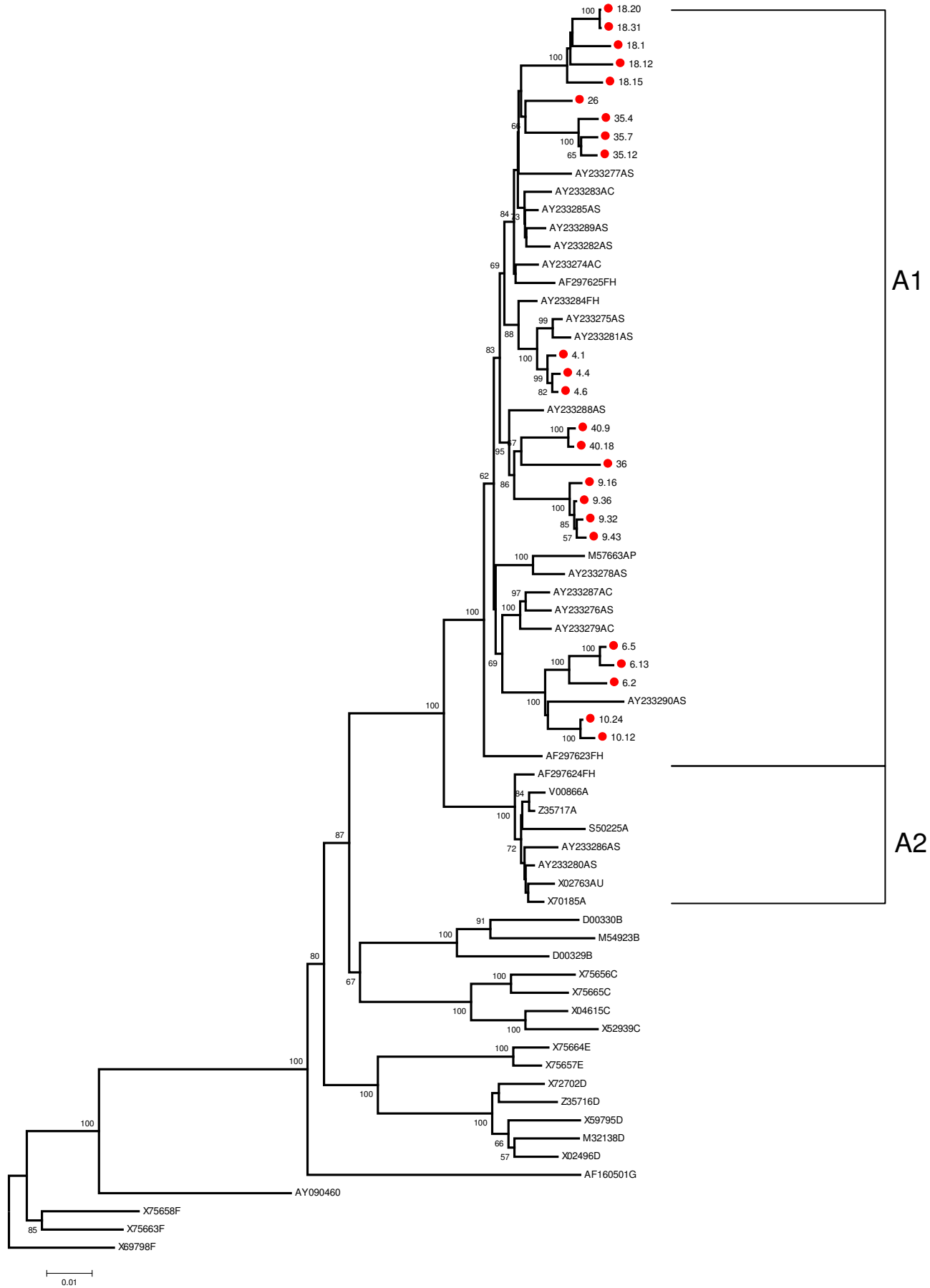


Figure 3.1. Complete Genome.

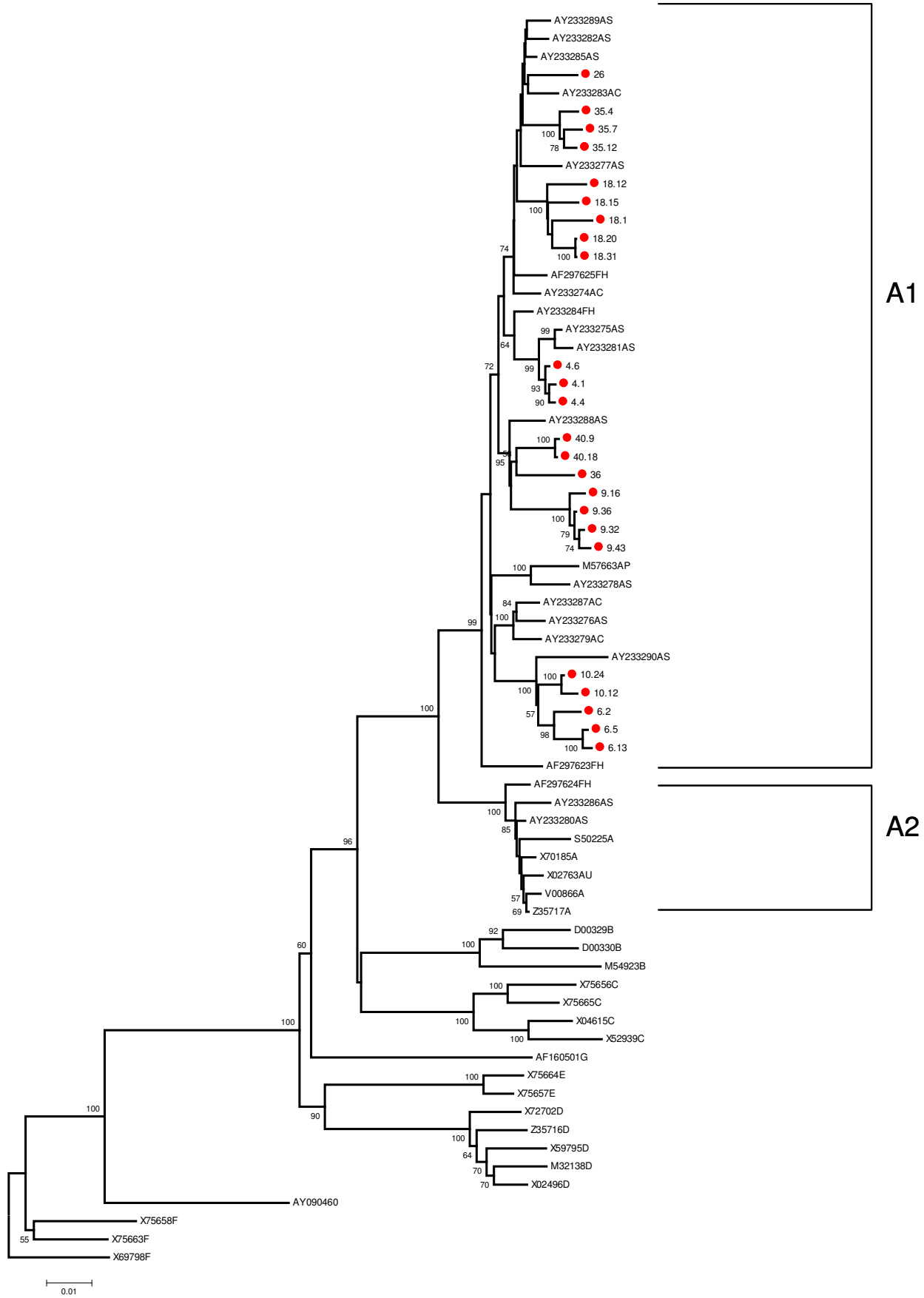


Figure 3.2. Polymerase Gene.

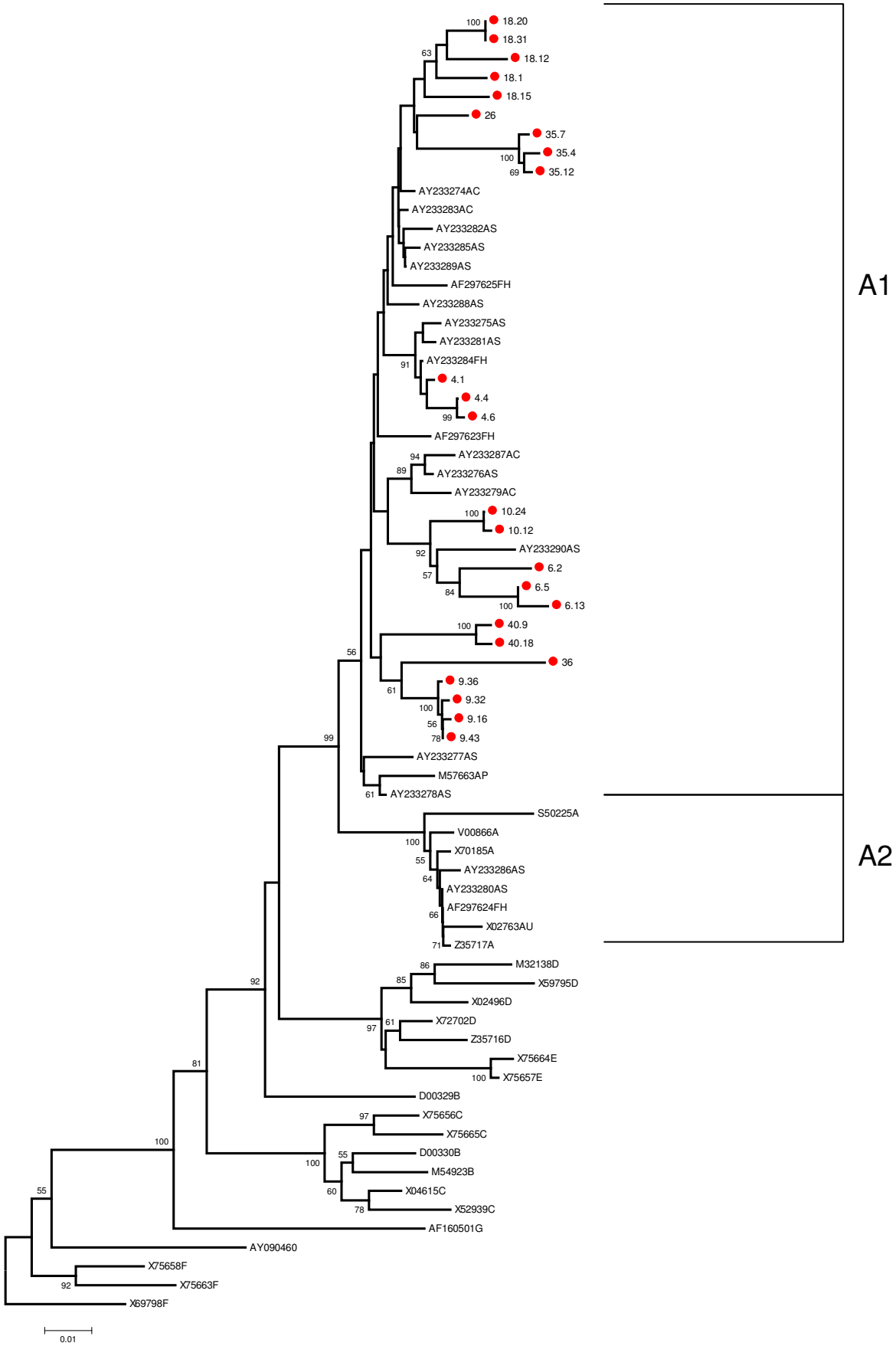


Figure 3.3. pre-C/Core Gene.

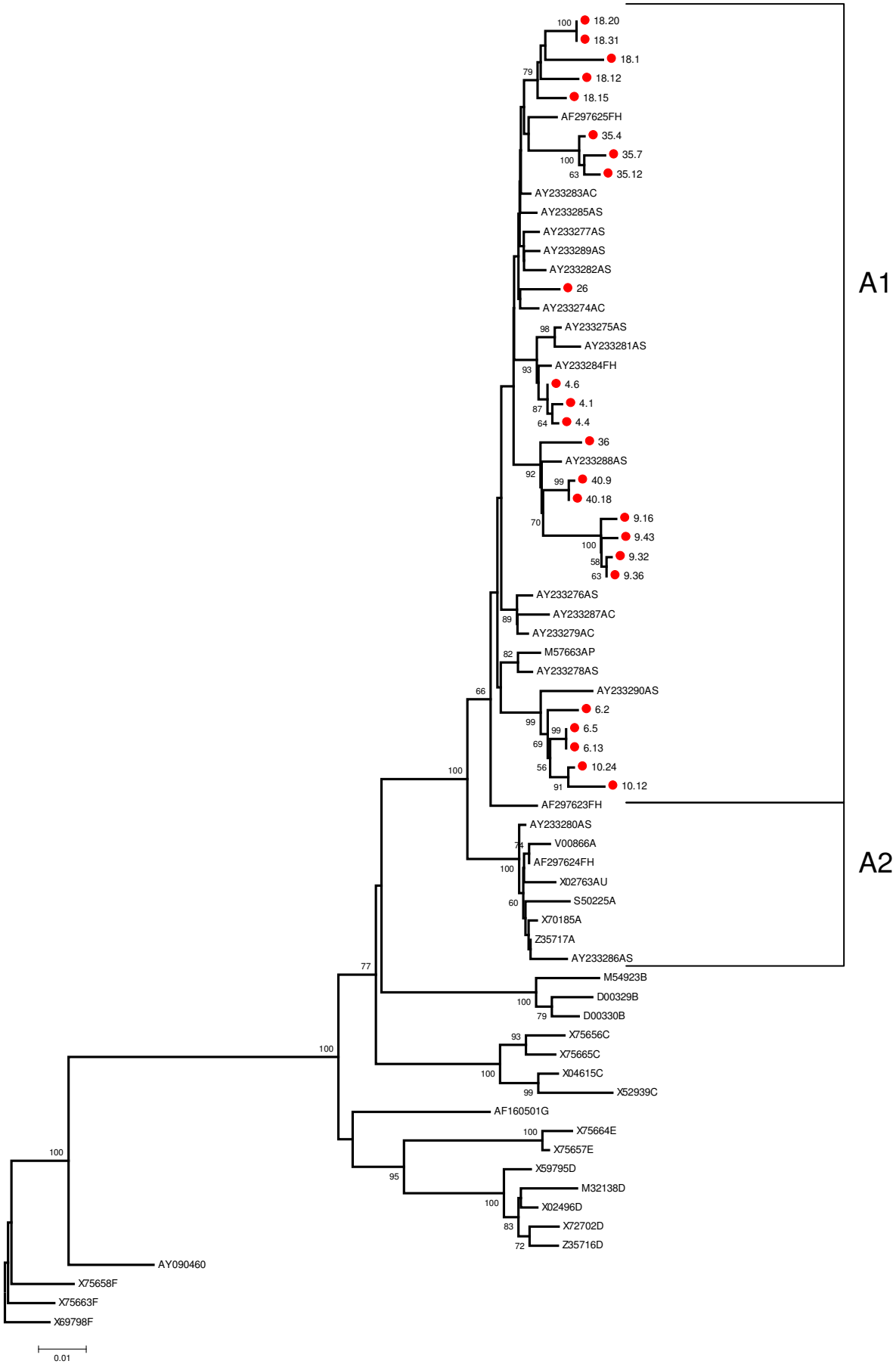


Figure 3.4. Surface Gene.

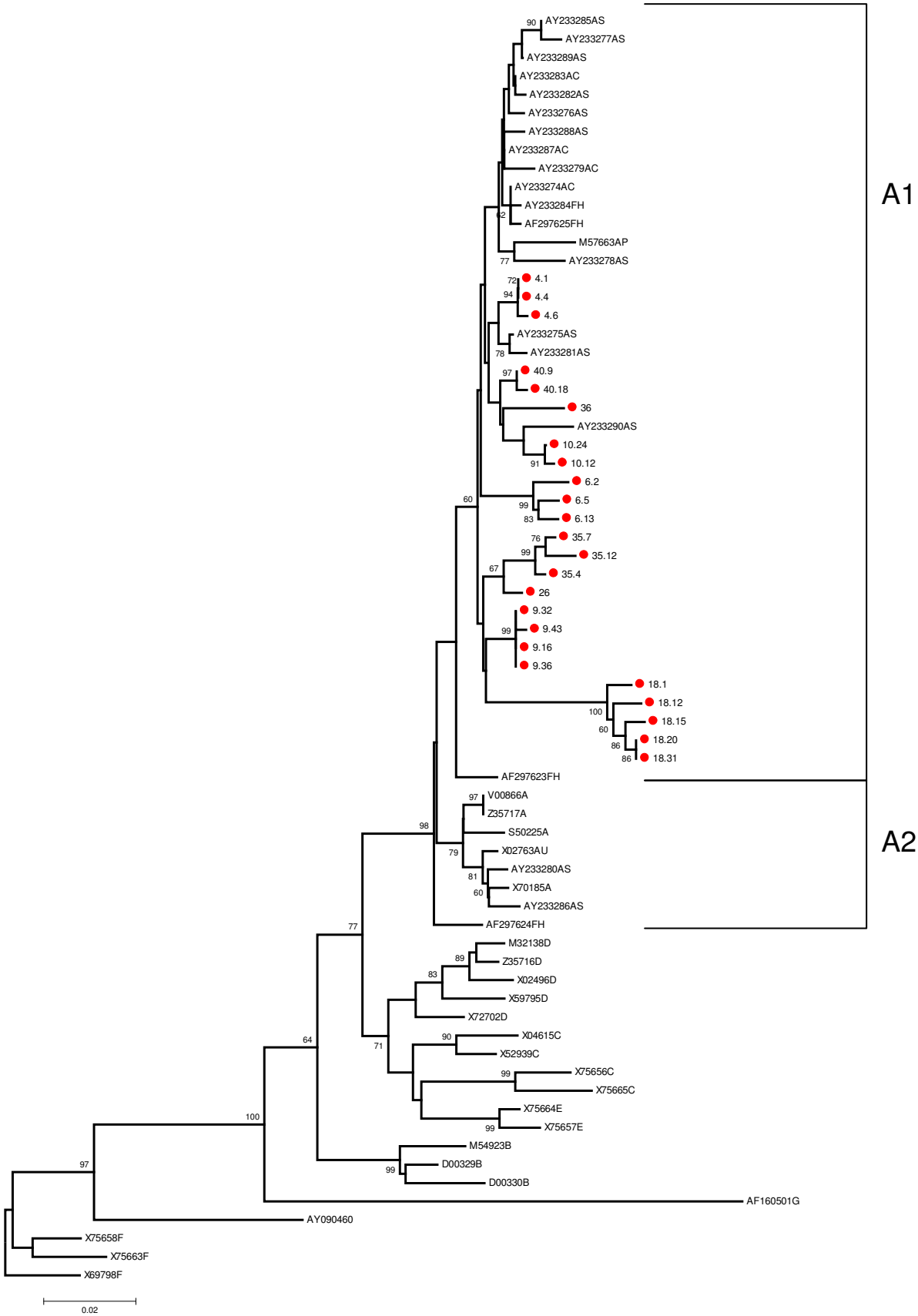


Figure 3.5. X gene.

Figures 3.1-3.5

Phylogenetic relationship of HBV isolates cloned and sequenced in the present study (red) to representative full-length sequences belonging to genotypes A to H, obtained from GenBank, established using neighbour-joining. Bootstrap statistical analysis was performed using 1 000 data sets and the numbers on the nodes indicate the percentage of occurrences (Felsenstein, 1995). Each sequence is designated by its accession number, type (A to H) and in the case of genotype A: AS = isolates from asymptomatic chronic carriers, AC = isolates from acute hepatitis patients and FH = isolates from fulminant hepatitis patients.

3.1.3. COMPARISON OF AMINO ACID SEQUENCES OF SUBGENOTYPE A1 HBV ISOLATES FROM HCC PATIENTS WITH THOSE FROM ASYMPTOMATIC CHRONIC CARRIERS, ACUTE HEPATITIS PATIENTS AND FULMINANT HEPATITIS PATIENTS

A comparison of variations in the amino acid residues of 24 HBV sequences derived from HCC patients in the present study relative to 17 sequences derived from GenBank is shown in figure 3.6. The amino acids found in subgenotype A1 only are shaded in grey. There is more variation in the pre-S1 region in the strains derived from the HCC patients compared to those from the asymptomatic chronic carriers and acute hepatitis patients. The HCC derived strains show variation at residues 67, 74, 86 and 91 in pre-S1. The clones derived from four of nine (4/9) HCC patients had an Ala at residue 86 compared to the isolates in 8/10 asymptomatic chronic carriers, 4/4 acute hepatitis patients and 2/2 fulminant hepatitis patients. Leu^{32} in pre-S2, Thr^{236}, Tyr^{269} and Gln^{334} in the spacer region were well conserved. Four of 9 HCC patients had HBV with Gly at residue 268 compared to 8/10 asymptomatic chronic carriers, 4/4 acute hepatitis patients and 2/2 fulminant hepatitis patients. HBV isolated from patients 6 and 40, had the amino acid pattern Ala^{86} in pre-S1 and Gly^{268} in the polymerase spacer region. HBV isolated from patients 18 and 35 had Gln^{338}. Four HBV clones from patient 18 had Gln and 1 with Lys at position 338. Ser^{146} and Ser^{147} in the X region characteristic of subgenotype A1 were also present in HBV isolated from HCC patients. All the clones of HBV isolated from patient 9 had Thr^{31} in the X region, whereas the HBV isolates from the other patients had alanine.

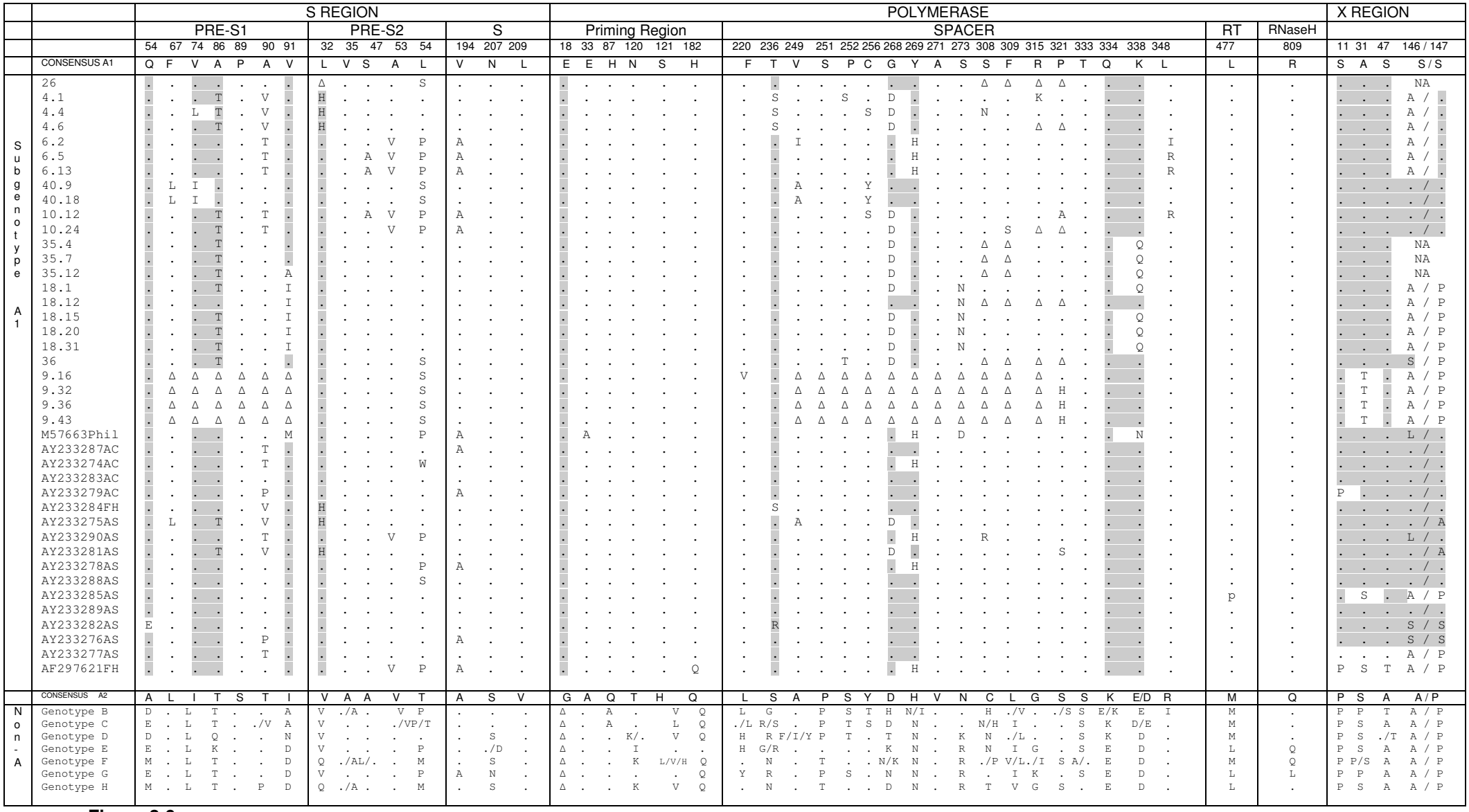

S REGION
PRE-S1
PRE-S2
S
POLYMERASE
Priming Region
SPACER
RT
RNaseH
X REGION
CONSENSUS A1
CONSENSUS A2
Subgenotype A1
Non-A
26
4.1
4.4
4.6
6.2
6.5
6.13
40.9
40.18
10.12
10.24
35.4
35.7
35.12
18.1
18.12
18.15
18.20
18.31
36
9.16
9.32
9.36
9.43
M57663Phil
AY233287AC
AY233274AC
AY233283AC
AY233279AC
AY233284FH
AY233275AS
AY233290AS
AY233281AS
AY233278AS
AY233288AS
AY233285AS
AY233289AS
AY233282AS
AY233276AS
AY233277AS
AF297621FH
Genotype B
Genotype C
Genotype D
Genotype E
Genotype F
Genotype G
Genotype H

Figure 3.6.

Figure 3.6. Comparison of amino acid residues of S, polymerase and X-ORFs of subgenotype A1 isolates from HCC patients, asymptomatic chronic carriers, acute hepatitis patients and fulminant hepatitis patients. Dots indicate amino acid identity with the consensus sequence, Δ an amino acid deletion. Amino acid residue 1 in the pre-S1 refers to the first amino acid of genotype A sequences. All the HCC isolates were sequenced in this study and all the other sequences were obtained from GenBank. For the non-A genotypes the representative amino acid was determined by aligning 10 sequences of each genotype from geographically distinct regions and the consensus amino acid deduced if it occurred in 60% or more of the sequences. Amino acids found only in subgenotype A1 and in other non-A genotypes (B-H) but not subgenotype A2 are in shaded grey. The consensus alignment residue was obtained using an alignment containing 24 isolates sequenced in this study and 49 GenBank sequences. V= A/G/C; W=T/C/G; X= A/T/C; Y=T/A/C and Z=G/C/T ranging from the most prevalent nucleotide to the least. Acute hepatitis (AC), fulminant hepatitis (FH) and asymptomatic chronic carrier (AS).

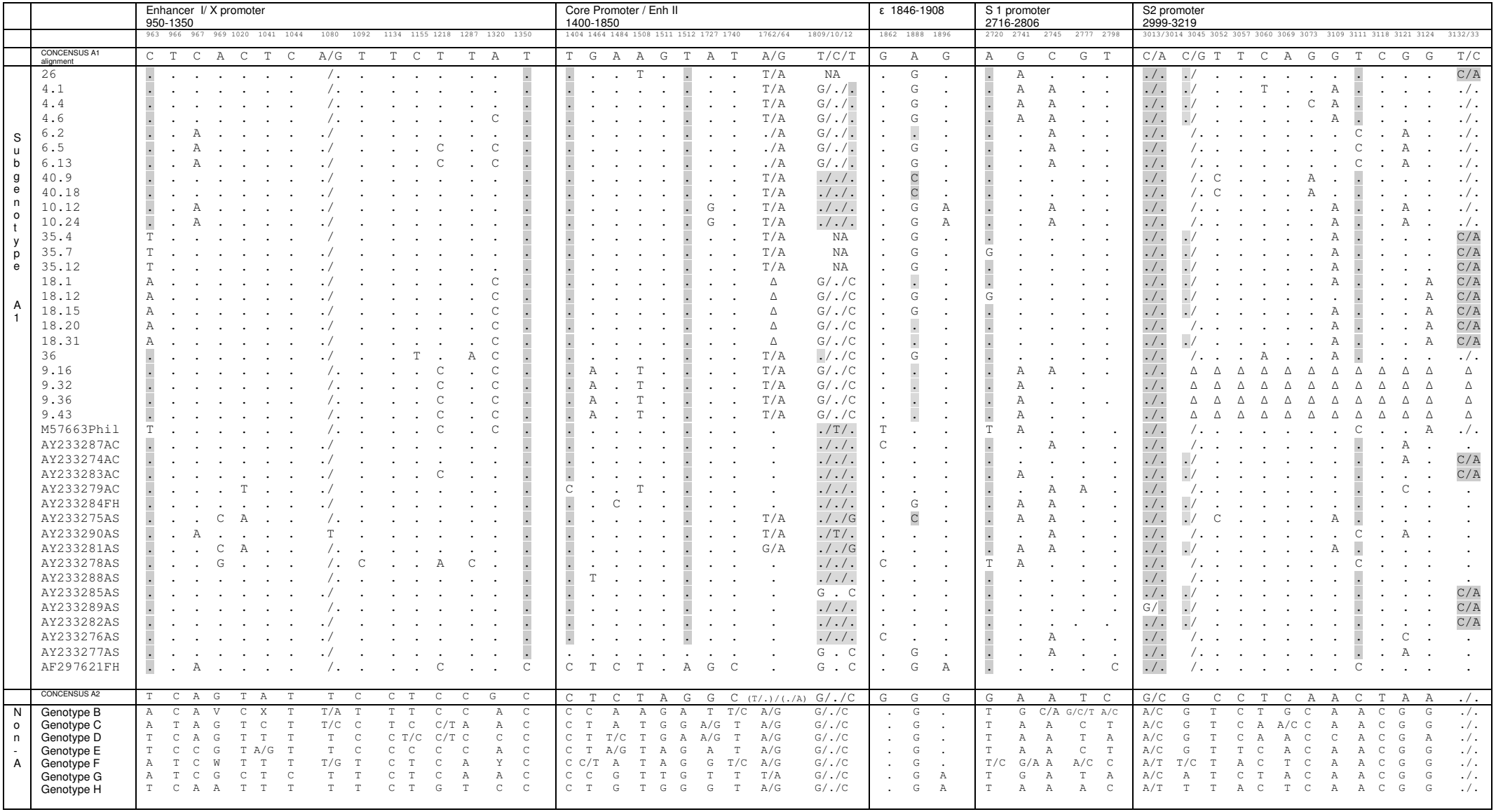

Enhancer I/ X promoter
950-1350
Core Promoter / Enh II
1400-1850
ε 1846-1908
S 1 promoter
2716-2806
S2 promoter
2999-3219
CONCENSUS A1 alignment
Subgenotype A1
CONCENSUS A2
Non-A
Genotype B
Genotype C
Genotype D
Genotype E
Genotype F
Genotype G
Genotype H
26
4.1
4.4
4.6
6.2
6.5
6.13
40.9
40.18
10.12
10.24
35.4
35.7
35.12
18.1
18.12
18.15
18.20
18.31
36
9.16
9.32
9.36
9.43
M57663Phil
AY233287AC
AY233274AC
AY233283AC
AY233279AC
AY233284FH
AY233275AS
AY233290AS
AY233281AS
AY233278AS
AY233288AS
AY233285AS
AY233289AS
AY233282AS
AY233276AS
AY233277AS
AF297621FH

Figure 3.7.

Figure 3.7. Comparison of nucleic acid sequences of the *cis*-acting elements of subgenotype A1 isolates from HCCs, asymptomatic chronic carriers and acute hepatitis patients. Dots indicate nucleotide identity. Nucleotide 1 denotes the nucleotide position of hepatitis B virus *adw* genome (GenBank accession AY233276) where the *Eco*RI cleavage site is at position 1. HCC isolates were all sequenced in the present study and all other sequences were obtained from GenBank. For the non-A genotypes the representative nucleic acid was determined by aligning 10 sequences of each genotype from geographically distinct regions and the consensus amino acid deduced if it occurred in 60% or more of the sequences. Amino acids found only in subgenotype A1 and in other non-A genotypes (B-H) but not subgenotype A2 are in shaded grey. Acute hepatitis (AC), fulminant hepatitis (FH) and asymptomatic chronic carrier (AS).

3.1.4. COMPARISON OF NUCLEOTIDE SEQUENCES OF SUBGENOTYPE A1 HBV ISOLATED FROM HCC PATIENTS WITH ISOLATES FROM ASYMPTOMATIC CHRONIC CARRIERS, ACUTE AND FULMINANT HEPATITIS PATIENTS BELONGING TO

Figure 3.7 illustrates the comparison of nucleotide sequences of the *cis*-acting elements of 24 HBV strains from HCC patients to 17 sequences derived from GenBank. Nucleotides found predominantly in subgenotype A1 but not in other genotypes are shaded in grey. Five of 9 HCC patients had strains of HBV with 1320C, whereas all the other patient groups had strains with 1320A. Patient 6 had 1320C in 2 clones of HBV and 1320A in 1 clone. Seven of the nine HCC patients had HBV with 1762A→T and eight of nine had 1764G→A. The 1809/1812 variation in the Kozak sequences, characteristic of subgenotype A1 was verified in the present study. Nucleotide 1888 in the precore region showed considerable variation in the HBV strains from HCC patients: 5/9 had 1888G only, 1/9 had 1888A only; 1/9 had 1888C and 2/9 had a mixture of clones with 1888G/1888A. Patients 6 and 18 had clones of HBV with either 1888G or 1888A. Both HBV clones derived from patient 10 had 1896A and 1858T (Appendix B figure B1). The majority of isolates derived from asymptomatic carriers had 1888A (8/10). Only 1 ASC carrier had HBV with 1888G. All four acute hepatitis patients had HBV with 1888A and the fulminant hepatitis patients had strains with 1888G. 3132T→C and 3133C→A in the S2 promoter previously shown to be characteristic of subgenotype A1 HBV was confirmed in the present study (figure 3.7).

Table 3.1.3. Mutations within the *cis*-acting regulatory elements found in subgenotype A1 HCC patients.

cis-acting element	Mutation/variation	Functional element affected
Enhancer I/X promoter	963 A/T→C	5' modulator element
	1092C→T	IRE
	1155T→C	RFX1
Core promoter	1404C→T	NRE-γ
	1464C/T→G	NRE-β
	1512A/G→T	NRE-β
	1809G→T, 1812C→T/G	precore Kozak Sequence*
ε	1888G→A	encapsidation signal
S1 promoter	2720T/G→A	HNF-1 binding site
S2 promoter	(3013A/G→C, 3014C/T→A)	NF1 transcription factor binding site
	3045T/A/G→C	region A
	3109A/C→G, 3111A/C→T	region E
	(3132T→C, 3133C→A)	region F

Numbering of nucleic acids according to genotype A [AY233276] and the functional domains according to (Moolla et al., 2002)
*Shown to reduce HBeAg translation by ribosomal leaky scanning mechanism (Ahn et al., 2003).
Functional domain according to (Garcia et al., 1993; Nakao et al., 1999; Moolla et al., 2002).

Table 3.1.4. Core promoter and precore mutations occurring in complete genomes of HCC patients in the current study (see Appendix B figure B1.nt 1753, 1766 & 1768).

HCC Isolate	1753 T→ C	1762 A→T	1764 G→A	1766 C→T	1768 T→A	1809 G→T	1812 C→T
6.1	-	-	+	+	+	-	+
6.2	-	-	+	+	+	-	+
6.5	-	-	+	+	+	-	+
6.13	-	-	+	+	+	-	+
9.16	-	+	+	+	-	-	-
9.32	-	+	+	+	-	-	-
9.36	-	+	+	+	-	-	-
9.43	-	+	+	+	-	-	-
10.12	+	+	+	+	-	+	+
10.17	+	+	+	+	-	+	+
10.24	+	+	+	+	-	+	+
36	+	+	+	+	+	+	-
40.9	+	+	+	-	-	+	+
40.18	+	+	+	-	-	+	+

3.2 SIMULTANEOUS ALTERATIONS IN THE BASIC CORE PROMOTER, CORE GENE AND SURFACE GENE

3.2.1. PCR and CLONING

DNA was extracted from the serum of a 28 year old black African female with HBV-induced HCC (patient 18). Full-HBV genome amplification was performed and resulted in a product of ~3.2 kb (figure 3.8A).This amplicon was successfully cloned into pCR Script Amp SK+ vector. Thirty nine positive clones were selected and *Pvu*II restriction digest revealed five clones (18.1, 18.12, 18.15, 18.20 and 18.31) with inserts of approximately 3.2 kb (figure 3.8B). The two *Pvu*II restriction patterns revealed in figure 3.8B already indicates a degree heterogeneity between these clones. The remaining clones had inserts of 1 kb or less. Initial sequence analysis revealed that all five clones were positive for HBV DNA. Full genome sequence analysis was carried out on clones 18.1, 18.12, 18.20 and 18.31. Phylogenetic analyses showed that all the isolates from patient 18 belonged to subgenotype A1 (section 3.1 figure 3.5).

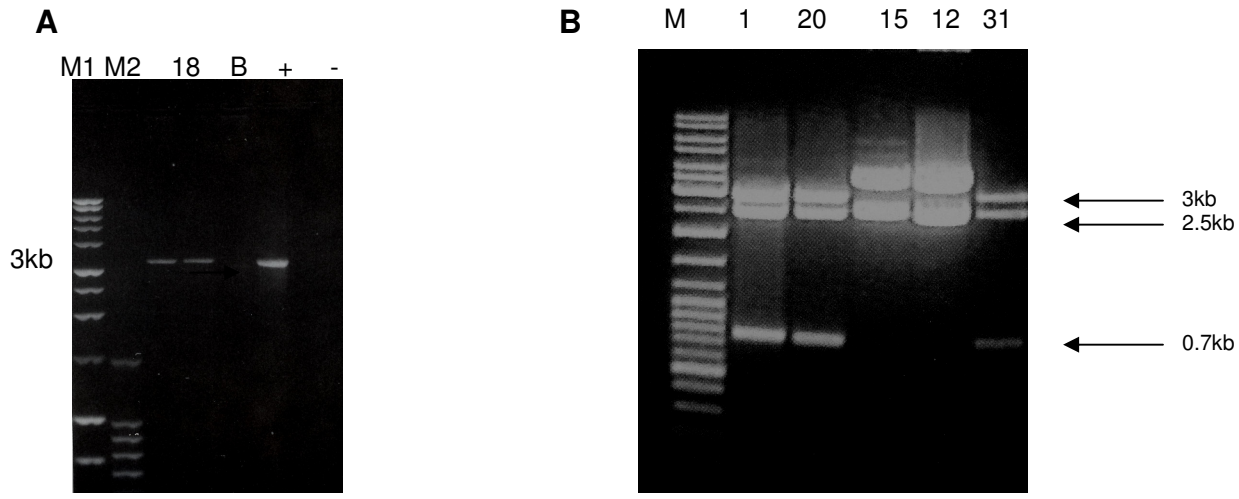


Figure 3.8.PCR amplicons and plasmid DNA restriction digest resolved on 1% agarose gels stained with ethidium bromide A) Full length genome amplicons for HCC 18 in duplicate, positive (+) and negative (-) controls are included. Amplicons were blunt end ligated into pCR Script Amp SK (+) B) *Pvu*II restriction digestion products of plasmid DNA for clones 18.1, 18.20, 18.15, 18.12 & 18.31. Clones 18.12 &18.15 have an internal *Pvu*II restriction site (CAGCTG) at nt 1302-1307, resulting in three fragments of ~0.7 kb, ~2.5 kb (vector) and ~3 kb. M1 1 kb DNA ladder, M2 100 bp DNA ladder and M O'Generuler DNA ladder mix.

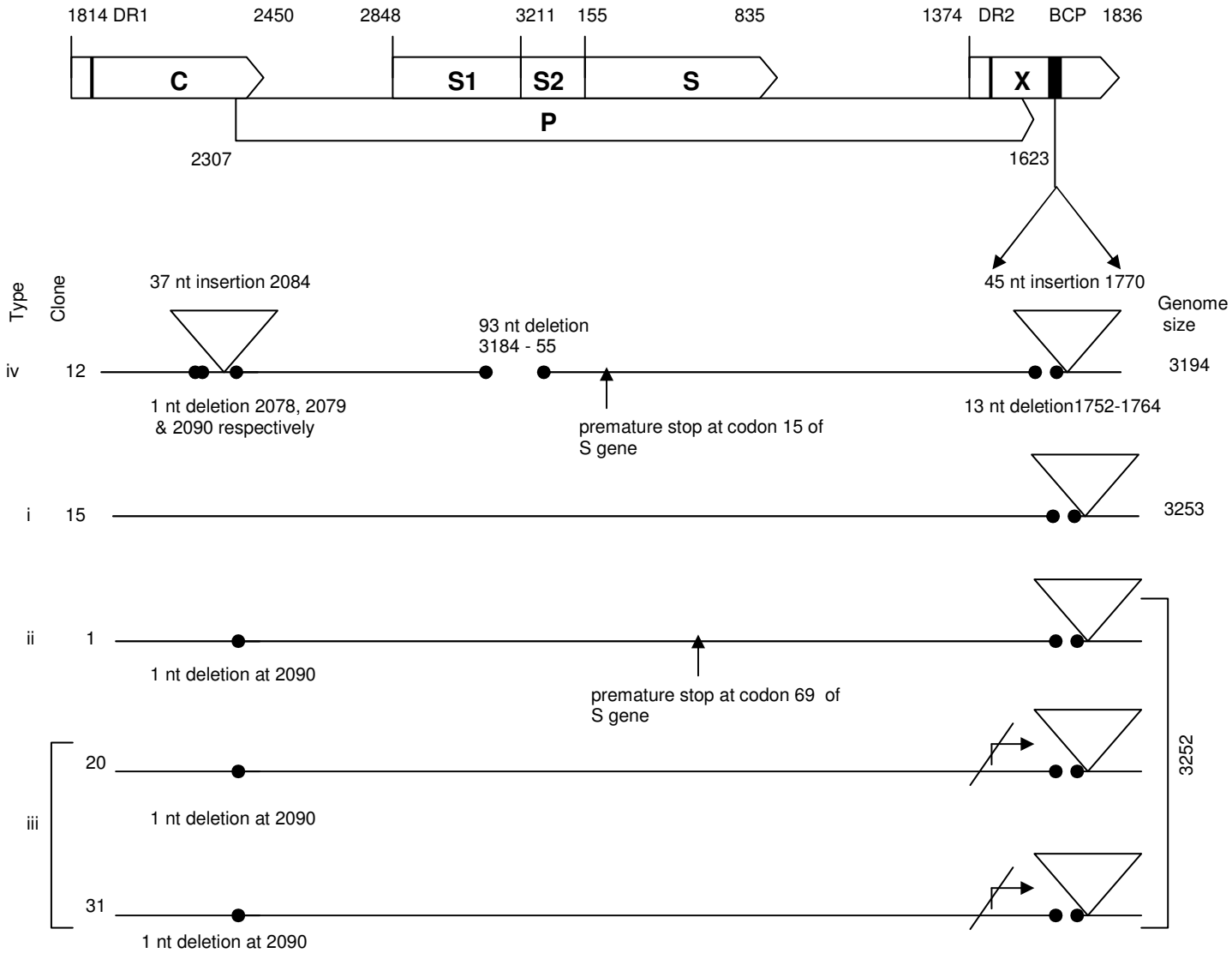


Figure 3.9. Full-genome structure of five HCC 18 complete genome HBV clones determined by sequence analysis. At the top is a schematic representation of the genome structure with four overlapping reading frames and the relevant *cis*-elements. Direct repeat 1 (DR1), basic core promoter (BCP). Insertions are indicated by the inverted triangle, deletions by dots, the X-gene initiation codon mutation is indicated by the cancelled arrow and S gene stop codons are indicated by arrows.

3.2.2. FULL GENOME STRUCTURE

Four types of mutant genomes were identified in patient 18 (figure 3.9).

- Type i (clone 18.15) contained a 13 bp deletion and a 45 bp insertion in the BCP region.
- Type ii (clone 18.1) contained the identical CP mutation as type i as well as a single nt deletion at nt 2090.
- Type iii (18.20 & 18.31) was similar to type ii but also contained an X ORF initiation codon mutation.
- Type iv (18.12) represented the most complex mutant of this set. Type iv contained all the changes in type ii as well as a 2 bp deletion at nt 2078 - 2079; a 37 bp insertion at nt 2084 in the core gene and a 93 nt deletion at nt 3184 - 55 overlapping the pre-S1/pre-S2 gene.

3.2.3. STRUCTURE OF THE BCP MUTATIONS

3.2.3.1. DELETION, INSERTION AND DUPLICATION

These mutations occurred within the BCP (nt 1741 - 1849). All five clones had a 13 nucleotide deletion at nucleotide position 1752-1764 and a 45 base pair insertion at nucleotide position 1770 (figure 3.9). The deletion site has some homology to the human topoisomerase I recognition site (5'AAAAAGACTT$^{\downarrow}$AGAAAAATTTTT3') (Konopka, 1988; Stewart et al., 1998) making it susceptible to recombination events (figure 3.10A). The insertion created two identical direct repeats of 27 nucleotides each (figure 3.10C). Analysis of the repeats revealed that they are made up of a combination of sequences homologous to 5' and 3' sequences adjacent to the site of insertion, including nt 1766 – 1776, nt 1755 – 1762 and nt 1770 – 1779 in a partially overlapping pattern

as illustrated in figure 3.10C. The middle AGGTCTT motif is homologous to nt 1645-1650.

3.2.3.2. INITIATION SITES AND TA-RICH MOTIFS

The initiation sites for the pre-C mRNA nt 1785 -1786 and nt 1791-1797 and the pgRNA 1815 ± 5 appear unaffected by the mutations (figure 3.10 A & B.) The AT rich motifs TA1, TA2 and TA3, which are needed for the optimal transcription of the pre-C mRNA, have undergone considerable modification (figure 3.10 B & C). TA1 was completely abolished. Four additional TA3 repeats were generated, while TA2 was duplicated. TA4 remained intact leaving the 'CATA' pre-C mRNA initiation sequence and the pgRNA TATA – box unaffected. A duplication of the 5'TAAATATTA3' sequence has occurred.

3.2.3.3. TRANSCRIPTION FACTOR BINDING SITES

The deletion 1750 – 1762 disrupted several overlapping *cis*-elements to different degrees. These include Sp1, HNF4, TBF and liver enriched factor (LEF) binding sites (figure 3.11A). It is possible that these alterations, within the transcription binding sites, could affect the regulation of transcription, although it is impossible to intimate the extent from the present study.

While the deletion mutation has disrupted several binding sites, the combination of the BCP deletion of nt 1752-1764, the 45bp insertion at nt position 1770 and substitution (1766C→G) (Table 3.2.1) mutations introduced a potential HNF3 site. Figure 3.11D provides a comparison with the HNF3 consensus sequence The insertion mutation appears to have introduced two HNF1 binding sites as indicated in figure 3.11B.

3.2.4. X-GENE MUTATIONS AND PREDICTED PROTEIN EXPRESSION

The BCP mutations inevitably affect the overlapping X open reading frame and can therefore influence the expression and function of the X protein. A premature stop codon (TAG) at position 1760 or at 388 nt downstream from the X ORF start site, may result in a truncated X protein of 130 amino acids in all five clones as illustrated in figure 3.12. Amino acids Ile Arg Leu (IRL) are substituted with Cys Leu Tyr (CLY). The Kunitz-type serine protease inhibitor domain is completely lost as a result of the mutations in these genomes. The latter domain is reported to be important for the putative transactivation activity of the X protein *in vitro* (Arii et al., 1992). Clones 18.20 and 18.31 also exhibited Met to Leu mutation at position 1 of X, which abolished the initiation codon preventing the expression of X protein from these genomes.

3.2.5. CORE GENE MUTATIONS AND PREDICTED PROTEIN EXPRESSION

Three types of core gene mutations were identified between nt 2078 – 2084 (figure 3.13).

- Type i (18.15) represents the 'wild type' sequence,
- Type ii (18.1, 18.20 & 18.31) contains a single nucleotide deletion at 2090.
- Type iii (18.12) contains the single nucleotide deletion at 2090, a 2 bp deletion at nt 2078-2079 and a 37 bp insertion at nt 2084.

When determining the effect of these mutations on core/HBe antigen expression, we found that type i would express normal core/e antigens (figure 3.13B); type ii, the single nt deletion at 2090 caused a frameshift in the C ORF and resulted in core protein truncation at Asn 64 (figure 3.13B) . In the type iii core gene mutant the insertion introduced a TAA stop codon resulting protein truncation after Asn 63. Detailed analysis of the 37 nt insertion revealed that it was comprised of a

combination of sequences homologous to 5' and 3' sequences adjacent to the site of insertion as illustrated by figure 3.13A.

3.2.6. SURFACE GENE MUTATIONS

Mutant genome type iv, the most complex of the five genomes, consists of a deletion in the surface ORF (nt 3 184 to 55). The surface gene mutations will be further characterised in section 3.4 and discussed in section 4.4.

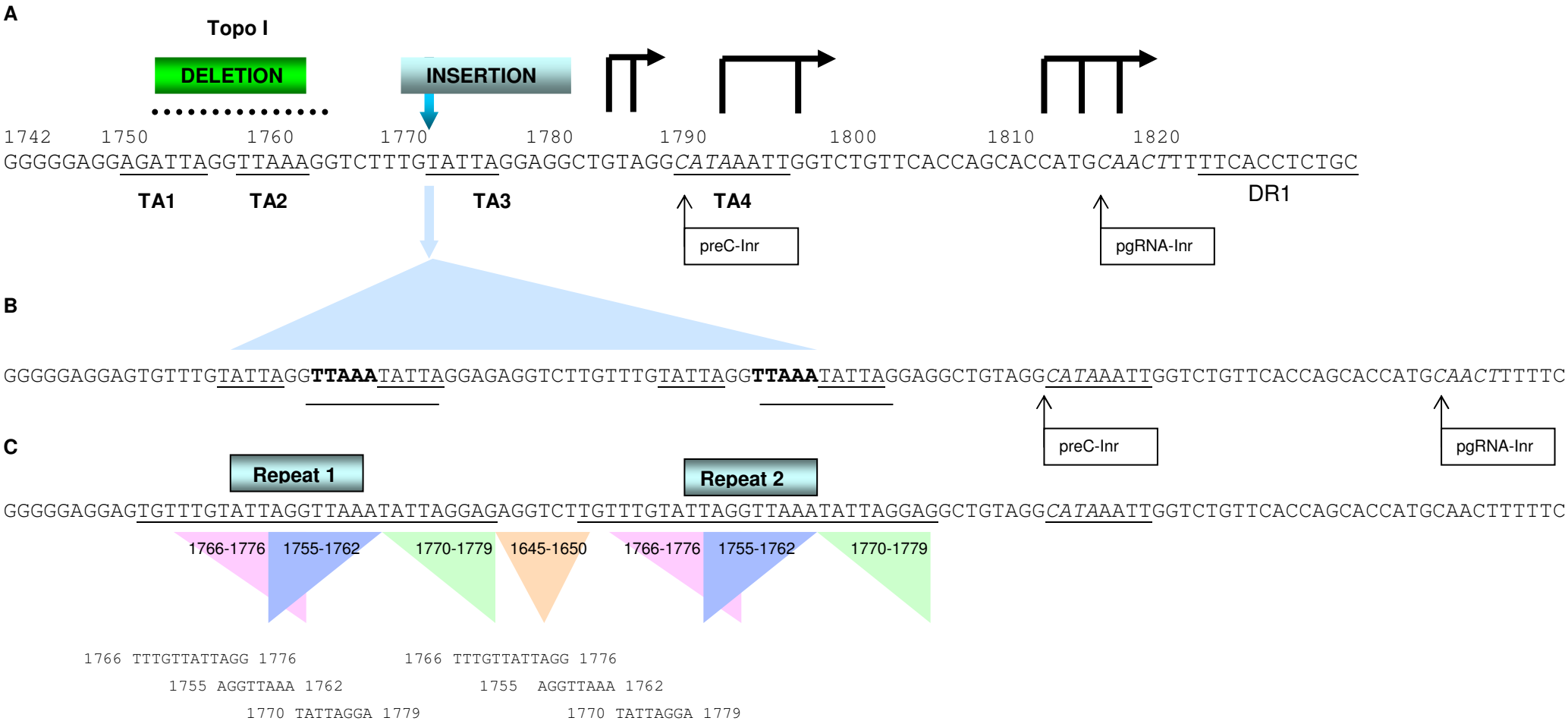
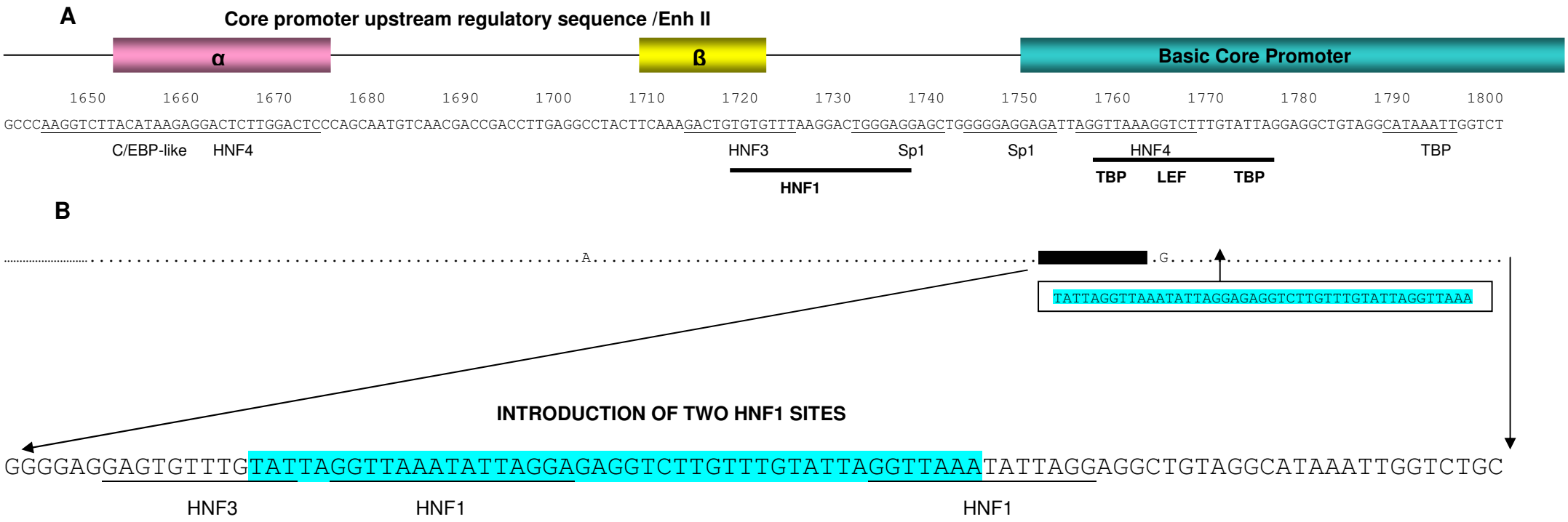

A
Topo I
DELETION
INSERTION
1742 1750 1760 1770 1780 1790 1800 1810 1820
GGGGGAGGAGATTAGGTTAAAGGTCTTTGTATTAGGAGGCTGTAGGCATAAATTGGTCTGTTCACCAGCACCATGCAACTTTTTCACCTCTGC
TA1
TA2
TA3
TA4
DR1
preC-Inr
pgRNA-Inr
B
GGGGGAGGAGTGTTTGTATTAGGTTAAATATTAGGAGAGGTCTTGTTTGTATTAGGTTAAATATTAGGAGGCTGTAGGCATAAATTGGTCTGTTCACCAGCACCATGCAACTTTTTC
preC-Inr
pgRNA-Inr
C
Repeat 1
Repeat 2
GGGGGAGGAGTGTTTGTATTAGGTTAAATATTAGGAGAGGTCTTGTTTGTATTAGGTTAAATATTAGGAGGCTGTAGGCATAAATTGGTCTGTTCACCAGCACCATGCAACTTTTTC
1766-1776
1755-1762
1770-1779
1645-1650
1766-1776
1755-1762
1770-1779
1766 TTTGTTATTAGG 1776
1755 AGGTTAAA 1762
1770 TATTAGGA 1779
1766 TTTGTTATTAGG 1776
1755 AGGTTAAA 1762
1770 TATTAGGA 1779

Figure 3.10.

Figure 3.10. Nucleotide sequence of the BCP (nt 1741-nt1 849) genotype A. (A) Initiation sites of the precore (pre-C) mRNA and pregenomic (pg) RNA are shown as horizontal arrows and initiators (Inr) are shown in italics. The BCP contains four AT-rich domains (TA1-TA4). The positions of the 13 nucleotide deletion and the 45 nucleotide insertion are indicated. B) Reorganization of the AT-rich domains in BCP for all five clones. TA2 motifs are shown in bold and a new motif TATTA and TTAAATATTA are underlined. C) Two 27 nucleotide repeats are underlined and are made up of a recombination of overlapping adjacent sequences, demonstrated by the overlapping triangles.



C

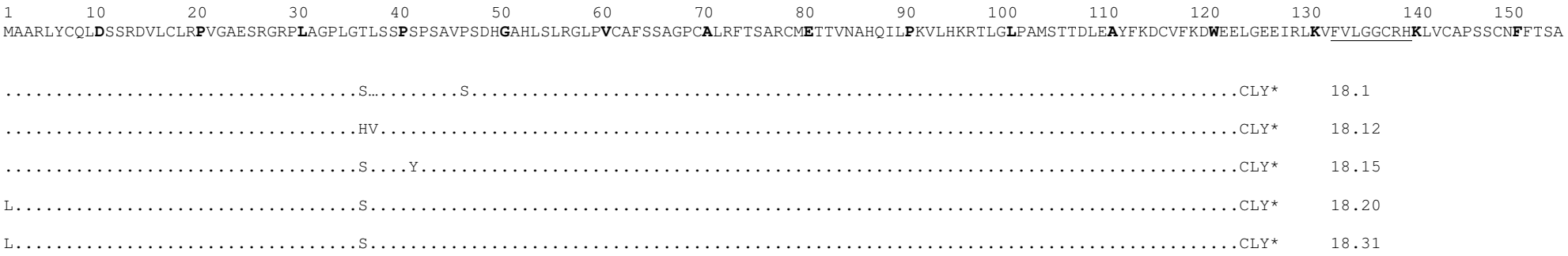
ENH I	TGT**GGTT**TAAGGACTG**GGA**G
New BCP	**GGTT**AAATATTA**GGA**
pre-S1 promoter	AGTTAATCATTACTT

D

patient	GAGTGTTTGTAT
	.. ..
HNF3 consensus	VAWTRTTKRYTY

Figure 3.11.

Figure 3.11. Nucleotide sequence of the core promoter showing transcription factor binding sites. A) CCAAT/enhancer binding protein (C/EBP) (Yuh and Ting, 1991; Lopez-Cabrera et al., 1990; Lopez-Cabrera et al., 1991), hepatocyte nuclear factor 1 (HNF1) (Wang et al., 1998), HNF 3 (Li et al., 1995; Raney et al., 1997), HNF 4 (Guo et al., 1993; Raney et al., 1997), liver-enriched factor (LEF) (Galibert et al., 1979; Lopez-Cabrera et al., 1990; Zhang and McLachlan, 1994; Buckwold et al., 1996; Buckwold et al., 1997), Sp 1 (Zhang et al., 1993) and the TATA- binding protein (Chen et al., 1995; Yu et al., 2005) are shown. B) Nucleotide alignment of mutant sequence representative of all five HCC 18 clones. The solid black line represents the deletion, the inserted sequence is indicated in the text box. The consequences of the mutation shows that two consecutive HNF1 sites are introduced. C) Comparison of HNF1 motifs and homologous nucleotides are underlined in bold. D) Comparison of HNF3 motifs and the dots indicate homology between nucleotides.



Fligure 3.12. Predicted X protein amino acid sequences expressed from mutants. Sequences were aligned to consensus sequence of the X protein. The consensus sequence for subgenotype A1 was generated using GenBank sequences listed in figure 3.22 A and B (M5766AP-AF297623) and BioEdit with a 55% majority at each position. Asterisks indicate stop codons. The Kunitz domain important for transactivation is underlined.

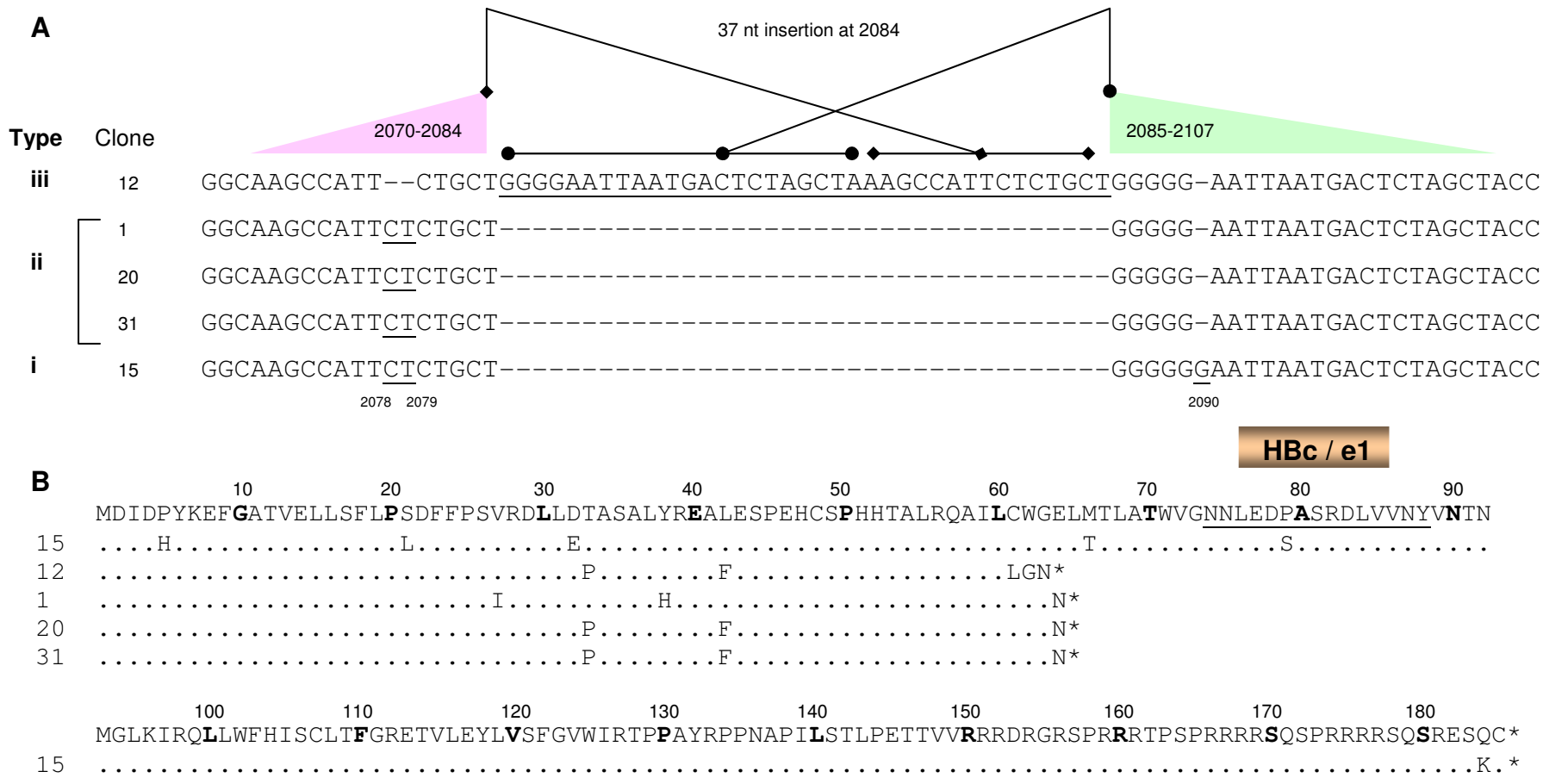

A
37 nt insertion at 2084
2070-2084
2085-2107
Type
Clone
iii
12
GGCAAGCCATT--CTGCTGGGGAATTAATGACTCTAGCTAAAGCCATTCTCTGCTGGGGG-AATTAATGACTCTAGCTACC
ii
1
GGCAAGCCATTCTCTGCT----------GGGGG-AATTAATGACTCTAGCTACC
20
GGCAAGCCATTCTCTGCT----------GGGGG-AATTAATGACTCTAGCTACC
31
GGCAAGCCATTCTCTGCT----------GGGGG-AATTAATGACTCTAGCTACC
i
15
GGCAAGCCATTCTCTGCT----------GGGGGGAATTAATGACTCTAGCTACC
2078 2079
2090
HBc / e1
B
10 20 30 40 50 60 70 80 90
MDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASALYREALESPEHCSPHHTALRQAILCWGELMTLATWVGNNLEDPASRDLVVNYVNTN
15 ....H.....L.....E.....T.....S.....
12 .....P.....F.....LGN*
1 .....I.....H.....N*
20 .....P.....F.....N*
31 .....P.....F.....N*
100 110 120 130 140 150 160 170 180
MGLKIRQLLWFHISCLTFGRETVLEYLVSFGVWIRTPPAYRPPNAPILSTLPETTVVRRRDRGRSPRRRTPSPRRRRSQSPRRRRSQSRESQC*
15 .....K.*

Figure 3.13.

Figure 3.13. Structure of the core gene mutation. A) Nucleotide sequence alignment of core gene sequence nt 2068-2110 of five HCC 18 clones. Clone 18.12 shows a 2 nt deletion 2078-2079 and a 37 nt insertion at position 2084. Four of the clones 18.12, 18.1, 18.20 & 18.31 have a single nt deletion at position 2090. The shaded triangles and lines indicate homology to adjacent sequences. B) Predicted core protein sequences. The consensus sequence for subgenotype A1 was generated using GenBank sequences listed in figure 3.22 A and B (M5766AP-AF297623) and BioEdit with a 55% majority at each position. The position of the major B-cell epitope of hepatitis B core antigen and HBeAg (HBc/e1) are shown above the sequences.

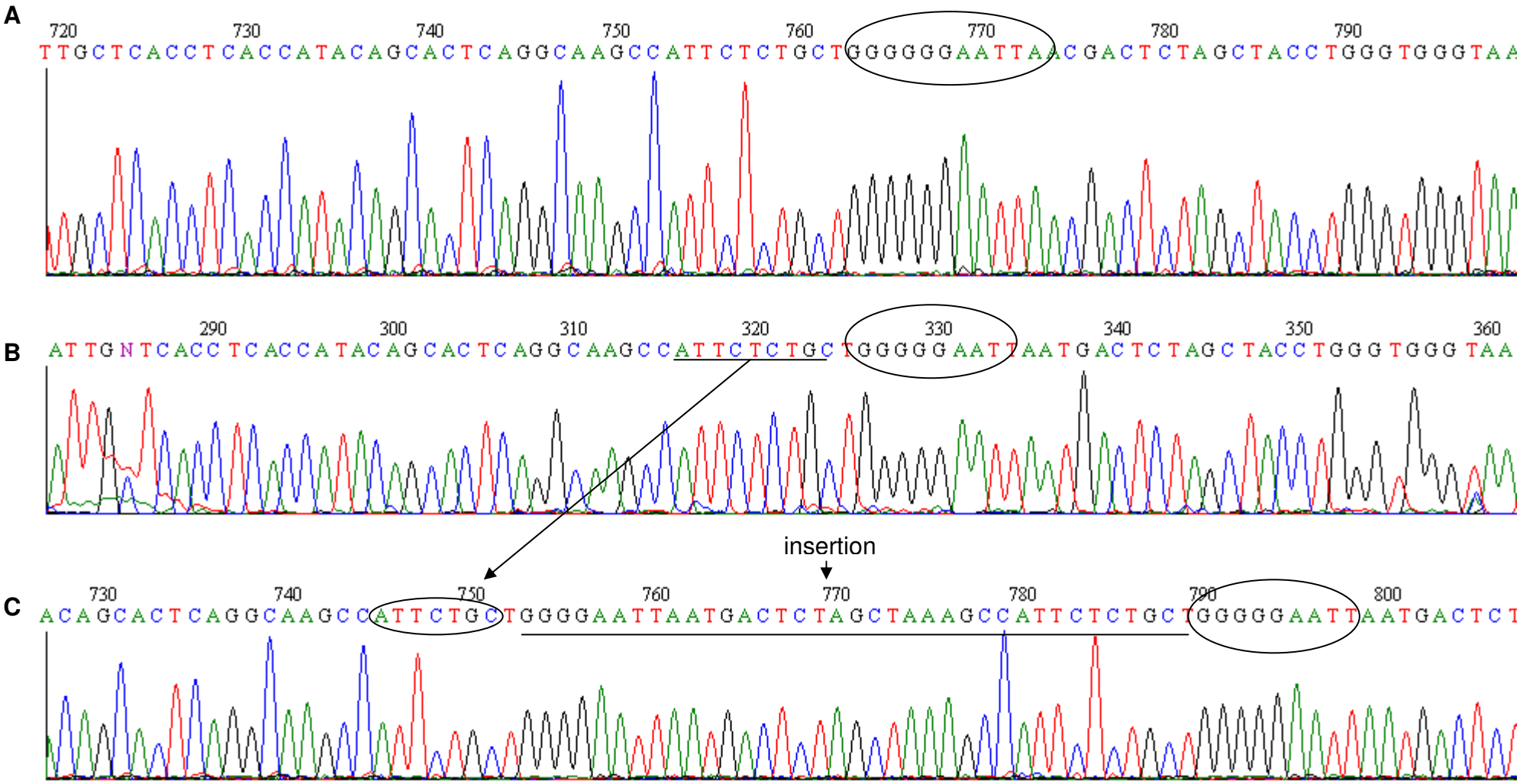

A
720
730
740
750
760
770
780
790
TTGCTCACCTCACCATACAGCACTCAGGCAAGCCATTCTCTGCTGGGGGGAATTAACGACTCTAGCTACCTGGGTGGGTAA
B
290
300
310
320
330
340
350
360
ATTGNTCACCTCACCATACAGCACTCAGGCAAGCCATTCTCTGCTGGGGGAATTAATGACTCTAGCTACCTGGGTGGGTAA
insertion
C
730
740
750
760
770
780
790
800
ACAGCACTCAGGCAAGCCATTCTGCTGGGGAATTAATGACTCTAGCTAAAGCCATTCTCTGCTGGGGGAATTAATGACTCT

Figure 3.14.

Figure 3.14. Chromatogram produced following automated sequencing of the core gene of HBV isolated from patient 18, visualized using Chromas version 1.45 (McCarthy, 1998) http://www.technelysium.com.au/chromas.html. A) Showing isolate 18.15 using primer 2509 (Table 2.2) which is representative of the wild type core gene sequence in the region. B) Showing isolate 18.31 using primer T3 (Table 2.2) showing a single nucleotide deletion. C) Shows isolate 18.12 using primer 2509, showing a 37 nt insert and deletion of nt 2078, 2078 and 2090.

Table 3.2.1. Mutations in the cloned HBV genomes detected by sequence analysis

	Cp/Enh II							C gene		Xgene	S1 promoter	S2 Promoter				Pre-S gene
	Deletions	Insertions	C 1766 >G	precore Kozak seq 1809/10/12 G / C / C	ε Signal G1888 >A	NRE γ	NRE β	Deletions	Insertions	X start	HNF1	NF1	region A	Region E	Region F	Pre-S1/S2 Δ
clone			nt position					nt postion						nt position		
18.12	1752-1764	45nt(1770)	C>G	G / C / C	-	*	*	2090; 2078-2079	37nt (2084)	intact	2720G	*	*	*	*	93nt (3184-55)
18.15	1752-1764	45nt(1770)	C>G	G / C / C	-	*	*	-	-	intact	*	*	*	3109A	*	-
18.1	1752-1764	45nt(1770)	C>G	G / C / C	G>A	*	*	2090	-	intact	*	*	*	3109A	*	-
18.20	1752-1764	45nt(1770)	C>G	G / C / C	G>A	*	*	2090	-	abolished	*	*	3045G	3109A	*	-
18.31	1752-1764	45nt(1770)	C>G	G / C / C	G>A	*	*	2090	-	abolished	*	*	*	3109A	*	-

* Except for the nt changes indicated in the table, all nt changes here are typical of subgenotype A1.NRE-γ-C1404T, NRE-β-C/T1464G & A/G1512T; HNF1-T/G2720A; NF1 A/G3013C, C/T3014A; Region A T/A/G3045C Region E C/A3109G, A/C3111T; Region F T3132C, C3133A

3.3. IDENTIFICATION AND MOLECULAR CHARACTERIZATION OF RARE HBV VARIANTS IN HEPATOCELLULAR CARCINOMA

Complete HBV genome amplification was performed on DNA extractions from forty black African patients with HBV-induced HCC (section 3.1). Thirty three patients (82.5%) were PCR positive (section 3.1). Twenty eight patients had a clear 3.2 kb fragment, while 13 of these also contained additional lower molecular weight fragments (0.5-3 kb). Five of the extracts that were PCR positive only yielded fragments smaller than 3.2 kb. The fragment size heterogeneity is illustrated in figures 3.15.A and 3.15.B. Patient 12 only had lower molecular weight bands present, while a mixture of fragments were present in patients 26 and 35.

Representative bands amplified from 10 patients are shown in figure 3.15 A. Extracts from patients 4, 6, 10, 13 and 14 were scored as positive in this experiment. Patient 6 and patient 10 had additional lower molecular weight fragments. These fragments were also of a lower intensity compared to the 3.2 kb fragment. The lower band intensity implies that these smaller genomes form a minor proportion of the virus population. This is supported by the fact that smaller fragments are amplified more efficiently than the larger 3.2 kb fragment even under stringent PCR conditions using *Taq/Pwo* in a 30 cycle PCR (Sommer et al., 1997). Smaller fragments were not present in the positive control (cloned HBV DNA template) indicating that these smaller genomes occur *in vivo* and were not artefacts of the PCR.

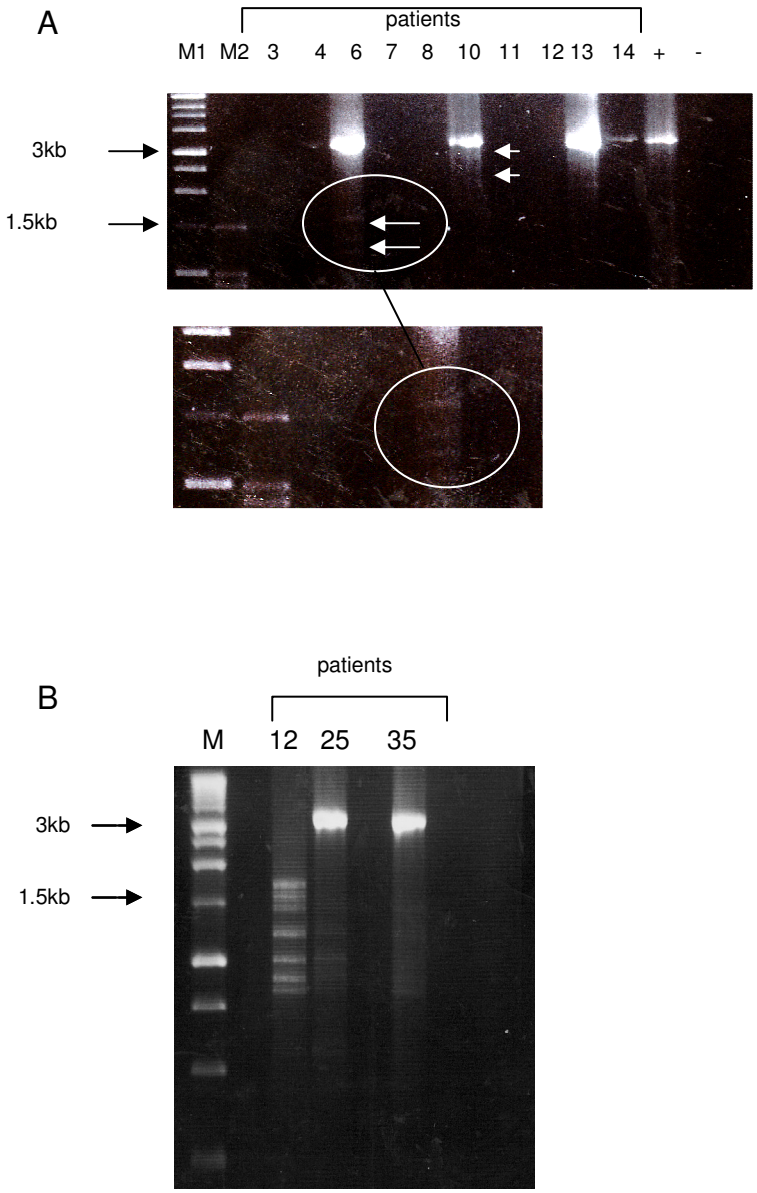


Figure 3.15. HBV full-length PCR amplicons from patient DNA resolved on 1% agarose gels stained with ethidium bromide. A) PCR using primers P1 and P2. M1 (1 kb DNA ladder), M2 (100 bp DNA ladder), patients, + (pSM2) and – (negative control) B) PCR using primers 1800 (+) and P2. M (1kb DNA ladder), patients, – (negative control).

3.3.1. HBV ISOLATED FROM PATIENT 6

The heterogeneous PCR products obtained from patient 6, a 40 year old black African male, were cloned into pCR Script Amp SK+ vector. Sixty one positive clones were selected. *Pvu*II restriction digestion revealed inserts of ~3.2 kb in 26 % of the clones, ~2 kb in 11 %, ~1.5 kb in 6 %, ~1kb in 5 % and less than 0.9kb in 52 %. Clones with various size inserts were selected for sequencing. Ambiguous sequences were excluded. A total of 4 clones were characterised completely. These include clone 6.1 with restriction pattern 6b and clones 6.2, 6.5 and 6.13 with restriction pattern 6c (figure 3.16).

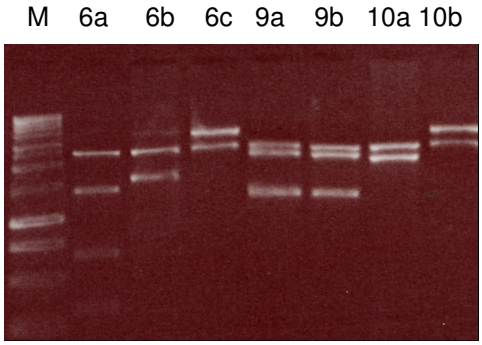


Figure 3.16. *Pvu*II restriction digestion products of cloned HBV resolved on a 1% agarose gel. M (1kb DNA ladder), clones 6a, 6b and 6c (patient 6), clones 9a and 9b (patient 9) and clones 10a and 10b (patient 10).

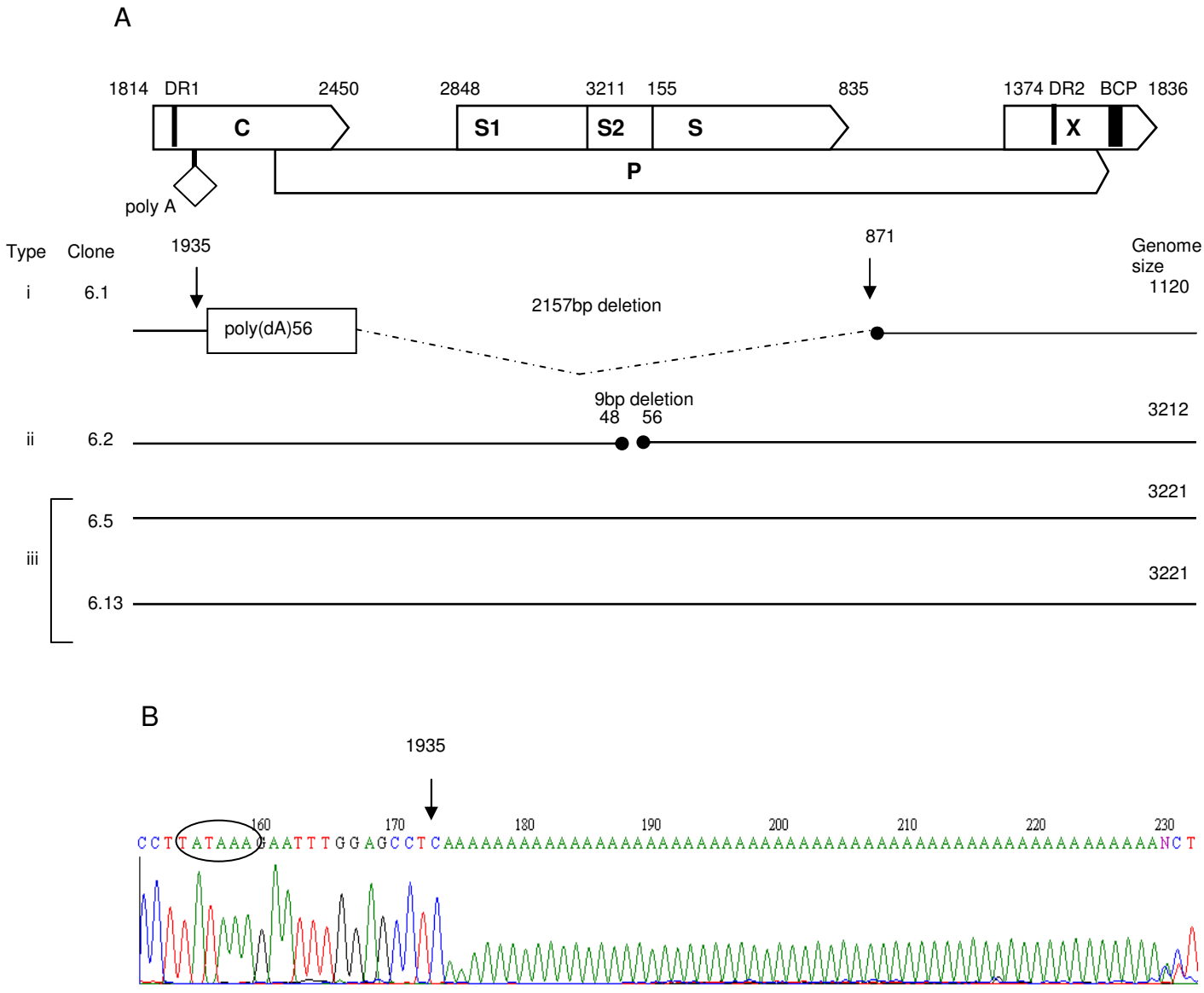

A
1814 DR1
2450
2848
3211
155
835
1374 DR2
BCP
1836
C
S1
S2
S
X
P
poly A
Type
Clone
1935
871
Genome size
i
6.1
1120
poly(dA)56
2157bp deletion
9bp deletion
48 56
ii
6.2
3212
3221
iii
6.5
3221
6.13
B
1935
160
170
180
190
200
210
220
230
CCTTATAAAAGAATTTGGAGCCTCAAAA

Figure 3.17 Full-genome structure of 6 clones derived from an HCC patient 6 determined using sequence analysis. A) At the top is a schematic presentation of the genome structure with four overlapping reading frames and the relevant *cis*-elements, including the basic core promoter (BCP), direct repeat 1 (DR1), direct repeat 2 (DR 2), polyadenylation signal (poly A). The poly (dA) tail is shown in the rectangular box, followed by a large deletion illustrated with the broken line and dots. B) Chromatogram of clone 6.1 HBV DNA sequence obtained using primer T3. The polyadenylation signal sequence (TATAAA) is circled and the 1935 nucleotide position is indicated.

3.3.1.1. GENOME HETEROGENEITY

Full genome sequence analysis revealed at least three types of mutant genomes in patient 6 (figure 3.17A).

- Type i (clone 6.1) contained an internal poly (dA) sequence as well as a large deletion resulting in a genome of 1 120 bp.
- Type ii (clone 6.2) contained a 9 bp deletion in the pre-S2 gene resulting in a genome of 3 212 bp.
- Type iii (clones 6.5 and 6.13) had no major deletions and have the expected wild type genome length of 3 221 bp.

3.3.1.2. STRUCTURE OF TYPE I GENOME

This genome contained an internal poly (dA) sequence of 56 nucleotides in length. The 5' – end of this poly (dA) sequence was located at position 1935, which is downstream to the processing/polyadenylation site (TATAAA) of the RNA pregenome (figure 3.17 B). A large deletion of 2 157 bp flanked the poly (dA) sequence with the 3' boundary at nucleotide position 871. The deletion site at nucleotide position 871 has not been described previously in this type of variant. Despite this large deletion, key elements required for viral replication and pregenomic RNA processing including the BCP, EN I and EN II, DR 2 and DR 2, the polyadenylation and encapsidation signals remained intact. It has been suggested that this genome was created by the coupled reverse transcription of the poly (A) tail of the RNA pregenome and partial DNA minus strand synthesis (Sommer et al., 1997).

3.3.1.3. PREDICTED PROTEIN SEQUENCE

The nucleotide sequences were translated using GeneDoc from the core initiation codon (nt 1901) and ending at the polymerase stop codon (nt 1623) for the poly (dA) variant as the deletion creates a new ORF and the polymerase start codon is lost. This protein has predicted size of 280 aa. The amino acids in grey are generated as a result of the mutation. The amino acids shaded in turquoise are homologous to the core protein aa 1-11 and region coloured in grey is homologous to the aa 599-844 of the viral polymerase.

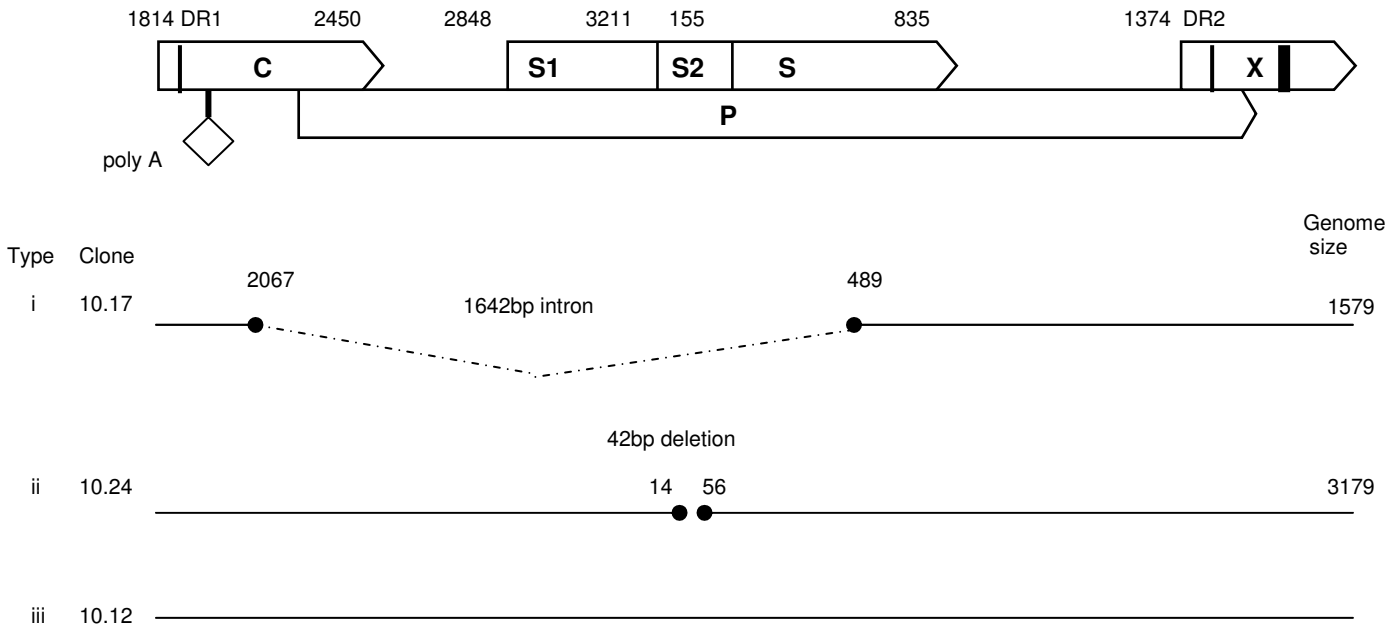
MDIDPYKEFGASKKKKKKKKKKKKKKKKKNFMGYVIGSWGTLPQDHIVQKI
KNCFRKLPVNRPIDWKVCQRIVGLLGFAAPFTQCGYPALMPLYASIQAKQAF
TFSPTYKAFLRNQYMNLYPVARQRPGLCQVFADATPTGWGLAIGHQRMRGTF
VAPLPIHTAELLAACFARSRSGANLIGTDNSVVLSRKYTSFPWLLGCTANWI
LRGTSFVYVPSALNPADDPSRGRLGLYRPLLRLPYRPTTGRTSLYAVSPSVP
SHLPDRVHFASPLHVAWRPP*280

Figure 3.18. Amino acid sequence of the recombinant protein generated for core-polymerase ORF of the poly (dA) variant. Genedoc was used to generate the sequence using the core initiation codon (nt 1901) and ending with the polymerase gene stop codon (nt 1623) .

3.3.2. HBV ISOLATED FROM PATIENT 10

The heterogeneous PCR products from HBV isolates obtained from patient 10, a 75 year old black African male, were cloned into pCR Script Amp SK+ vector. Thirty positive clones were selected. *Pvu*I restriction digestion revealed inserts of ~3.2 kb in 26% of the clones, ~2 kb in 20 % and 1.2-1.5 kb in 53 %. Clones with various size inserts were selected for sequencing. Ambiguous sequences were excluded. A total of 3 clones were characterised completely. These included clone 10.17 with banding pattern 10 a (figure 3.16) and 10.12 and 10.24 with banding pattern 10 b (figure 3.16).

A



B

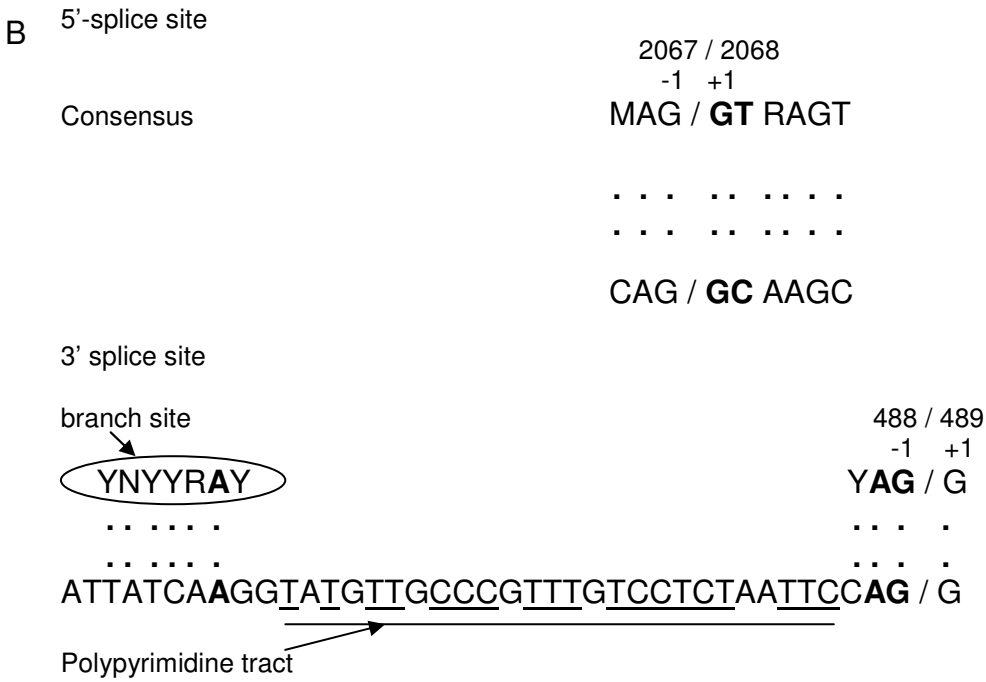


Figure 3.19

Figure 3.19 Molecular structure of three HBV genomes isolated from HCC patient 10 determined by complete sequence analysis. A) At the top is a schematic presentation of the genome structure with four overlapping reading frames and the relevant *cis* -elements including the basic core promoter (BCP), direct repeat 1 (DR1) and direct repeat 2 (DR 2). The large deletion is illustrated with the broken line and dots. B) Deletion boundary sequences of genome 10.17 were aligned to the consensus sequences for the 5' – splice site (MAG/GTRAGT), the 3' splice site (YAG/G), and the branch site (YNYYRAY) (Jackson, 1991).The branch site and polypyrimidine tract are indicated, non-negotiable nucleotides are shown in bold.

3.3.2.1. GENOME HETEROGENEITY

Complete genome sequence analysis revealed at least three types of HBV genomes in patient 10 (figure 3.19A).

- Type i contained an internal deletion of 1642 bp.
- Type ii contained a 42 bp deletion in the pre-S2 gene.
- Type iii had no major deletions and had the expected wild type genome length of 3221 bp.

3.3.2.2. STRUCTURE OF THE TYPE I GENOME

This genome contained an internal deletion of 1 642 bp at position nt 489 – nt 2067. The 5' boundary of the deletion revealed the presence of a donor splice site at position 2067/2068 and a common acceptor splice site at the 3' boundary at position 488/489 (figure 3.19B). This indicates that the internal sequence deleted is in fact an intron, which has been spliced from the pregenomic RNA. The remaining sequences between nucleotide 489 and 2067 also contained all the essential elements required for replication, RNA pregenomic processing, plus and minus DNA strand synthesis and nucleocapsid encapsidation.

3.3.2.3. PREDICTED PROTEIN SEQUENCE

A translated protein was generated with GeneDoc starting at the core initiation codon (nt 1901) and ending at the polymerase stop codon (nt 1623) for the splice variant. This protein has a predicted size of 433aa. The amino acids shaded in green are homologous to the core protein aa 1-56 and region coloured in grey are homologous to the aa 469-844 of the viral polymerase.

MDIDPYKEFGATAELLSFLPSDFFPSVRDLLDTASALYREALESPEH

CSPHHTALRINNNQYGTLQNLHDSCSRQLYVSLMLLYKTYGWKLHLY

SHPIILGFRKIPMGVGLSPFLLAQFTSAICSVVRRAFPHCLAFSYMD

DVVLGAKSVQHLESLYTAVTNFLLSLGIHLNHNKTKRWGYSLNFMGY

VIGSWGSLPQDHIVQKIKCCFRKLPVNRPIDWKVCQRIVGLLGFAAP

FTQCGYPALMPLYACIQAKQAFTFSPTYKAFLSKQYMNLYPVARQRP

GLCQVFADATPTGWGLAISHQRMRGTFVAPLPIHTAELLAACFARSR

SGAKLIGTDNSVVLSRKYTSFPWLLGCTANWILRGTSFVYVPSALNP

ADDPSRGRLGLYRPLLRLPYRPTTGRTSLYAVSPSVPSHLPVRVHFA

SPLHVAWRPP* 433

Figure 3.20 Amino acid sequence of the recombinant protein generated for core-polymerase ORF of the splice variant. Genedoc was used to generate the sequence using the core initiation codon (nt 1901) and ending with the polymerase gene stop codon (nt 1623).

The type i genomes described in patient 6 and patient 10 are similar, both contained the complete X ORF and truncated versions of core gene fused to other ORFs. The polymerase and surface coding regions were partially deleted, while the pre-S1 and pre-S2 were completely deleted. The surface gene regulatory domains, promoters S1 and S2 were completely abolished. It is also clear both patients displayed a heterogeneous population of genomes. Both patients contained a surface gene variant (type ii) and a full-length genome (type iii).

3.3.3. PHYLOGENETIC ANALYSIS

The X ORF was used to determine the genotype of variants 6.1 and 10.17 as this was the only uninterrupted gene in the variant. The entire X ORF of variants 6.1 and 10.17 were included into the X gene alignment carried out with 49 GenBank sequences and 25 HCC sequences. Both variants grouped within subgenotype A1 and within the corresponding patient group (Figure 3.21).

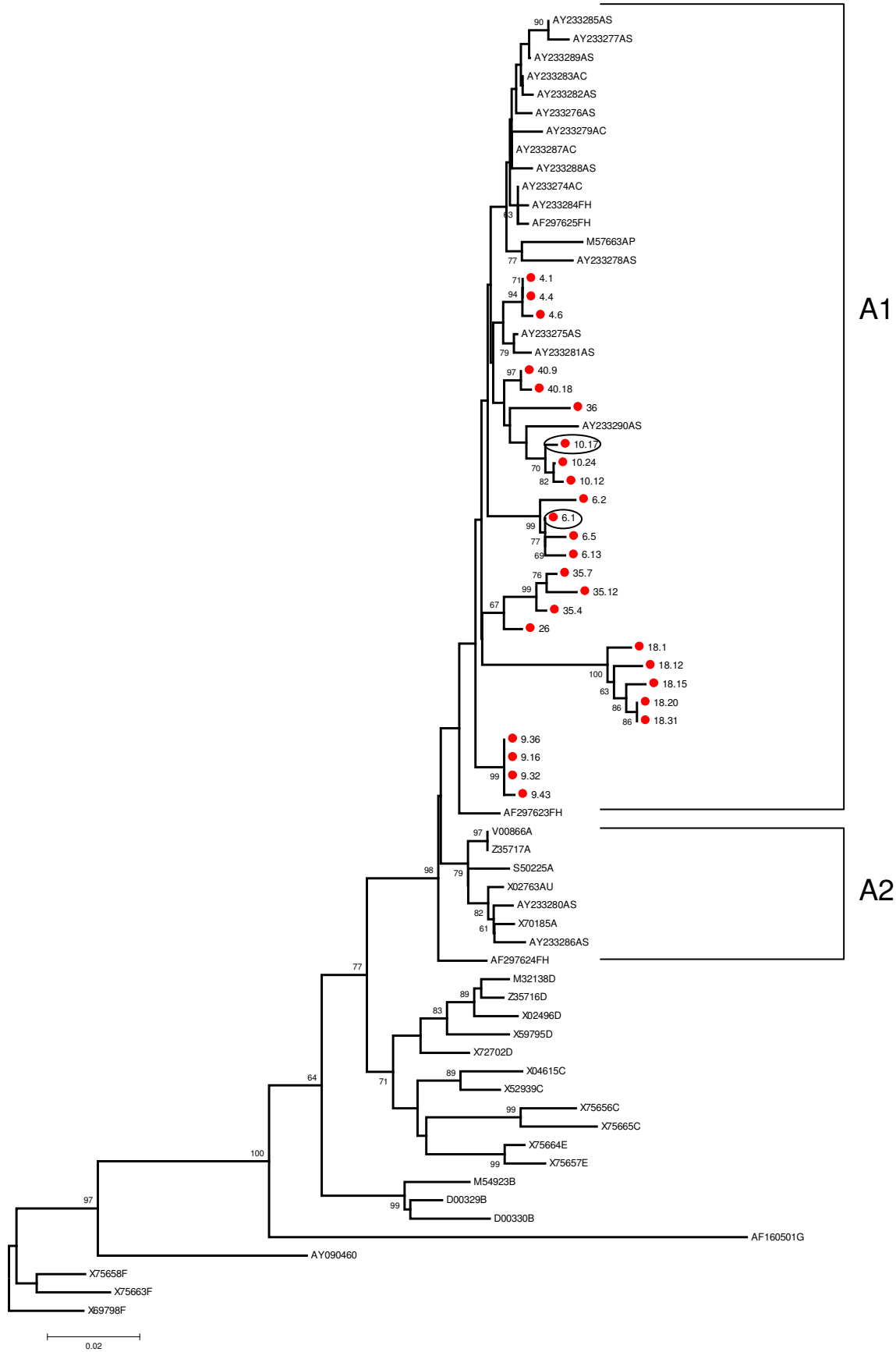


Figure 3.21. X gene

3.4 CHARACTERIZATION OF PRE-S MUTATIONS IN HBV ISOLATED FROM HCC PATIENTS

3.4.1. PRE-S MUTATIONS

The pre-S protein alignment (figure 3.22) compares 70 sequences (47 HBV strains from 18 HCC patients and 23 sequences from GenBank).

Two of the 47 isolates, 6.13 and 18.15, had a pre-S1 start codon mutation (M1V/K) and pre-S2 start codon mutations (M120A/T), respectively (figure 3.22 A & B) whereas 25% (12/47) the isolates had a pre-S2 start codon mutation (M120T/I/V/A) (Table 3.4.2). The hydrophobic Met is converted to the hydrophilic Thr in 5 isolates and to hydrophobic amino acids Ile, Val or Ala in 6 isolates. These results indicate that the pre-S2 domain is main target for mutation.

HBV isolated from thirteen of eighteen patients (72%) had a deletion in the pre-S gene in at least one isolate. 57% (27/47) of the all the isolates cloned demonstrated a deletion mutation in the pre-S region. 25% (12/47) of the deletions occurred in the pre-S1 region, 53% (25/47) in the pre-S2 and 19% (9/47) occurred across the pre-S1/pre-S2 boundary (Figure 3.22). The size of the deletions occurring within the pre-S region, ranged from 1 to 281 nucleotides (Table 3.22). The largest deletion occurred across the pre-S1/pre-S2 boundary. The 5' nucleotide position of the deletions appears to be more variable, when compared to the 3' nucleotide position. The most common deletion site occurred at nucleotide position 54. All the internal deletions were in frame.

Deletion mutations that partially or completely affect pre-S functional domains are listed in Table 3.4.1. The virion morphogenesis domain was mutated in 29% of the HCC isolates, Hsc70 interaction domain in 19%, cytosolic anchorage domain in

17% and the hepatocyte receptor domain in 2%. The region between residues 92-125 is relatively conserved compared to upstream stretches 1-10 and 80-91 and downstream region 130-155. This conserved domain coincides with residues 103-124 responsible for interacting with nucleocapsids. Residues 98, 99,102, 103, 105, 106, 113, 122 and 123 (indicated in boldface in line D of figure 3.22) are very well conserved and important for virion morphogenesis (Bruss, 1997). Several substitutions occur at the well conserved residues 122 and 123 (Table 3.4.2), and at less conserved positions and are also listed in Table 3.4.2.

46.8% (22/47) of the isolates from HCC patients contained deletions in the B cell epitope, while 51% (24/47) contained deletions affecting the T cell epitope. By disposing of protective epitopes, the virus presumably escapes detection by the host defence system, allowing for immune evasion and prolonged viral persistence.

Overall, 94% (17/18) of the HCC patients contained HBV pre-S gene mutations, these included deletions, initiation codon mutations and point mutations within the immune epitopes (figure 3.22).

3.4.2. PRE-S MUTATION PATTERN AND HETEROGENEITY

In the present study at least seven pre-S mutational patterns were recognized (figure 3.23).

- Type i contained a pre-S2 initiation codon mutation (M120T/I/V/A). This type of isolate would presumably express wild-type length LHBs and SHBs but no MHBs. This strain was found in combination with all other strains including the wild-type (figure 3.22).

- Type ii contained initiation codon mutations at both the pre-S1 (M1V/K) and pre-S2 start sites. Initiation of LHBs protein synthesis could presumably occur from the second Met at position 12 in the pre-S1 region, while no MHBs protein would be expressed.
- Type iii contained a pre-S2 initiation codon mutation and a pre-S2 internal deletion.
- Type iv contained a pre-S1 N-terminus deletion, which abolishes the first Met.
- Type v was similar to Type iv but had an additional pre-S2 deletion.
- Type vi consisted of a pre-S2 deletion and was the most common deletion mutant in the present study. It was found alone or together with other mutants. (figure 3.22B)
- Type vii consisted of a deletion across the pre-S1/pre-S2 boundary and resulted in the shortest LHBs (324-357aa). It was also found in combination with other mutants (patient 18) or alone (patient 9). A homogenous set as in patient 9, would imply the prevalence of that particular mutational pattern (variant).

Type ii occurred in two patients (6 and 18) with heterogeneous combinations. Patient 6 has pre-S types i, ii and iv as well as an internal poly (dA) variant described in section 3.3. Patient 18 had a combination of pre-S, Core and BCP mutations, described in detail in section 3.2. Type iii and type iv were isolated from the same patient (patient 4) and this pattern was not identified in any other patient in this study (figure 3.22). Patient 10 had a heterogeneous combination of isolates including pre-S type v, a spliced variant genome described in section 3.3.

3.4.3. C-TERMINALLY (3') TRUNCATED PRE-S/S GENE MUTANTS

Four types of 3' truncation mutants generated by a combination of pre-S gene deletion mutation and S gene stop codon mutations are illustrated in figure 3.24.

- Type i contained a pre-S1/pre-S2 deletion and an N-terminal nonsense mutation, which introduced a stop codon at residue 70 of MHB (or residue 15 of SHBs) these would designated $LHBs^{t70}$ (isolate 18.12).
- Type ii contained a pre-S1/pre-S2 deletion with a C-terminal stop codon at residue 254 of MHBs (or residue 199 of SHBs) designated $LHBs^{t254}$ (isolates 9.5, 9.16, 9.32, 9.36 and 9.43) or 271 of MHBs (or residue 216 of SHBs) and are designated $LHBs^{t271}$ (isolate 26).
- Type iii contained a stop codon at residue 124 (or residues 69 of SHBs) and are designated $LHBs^{t124}$/ $MHBs^{t124}$ (isolate 18.1).
- Type iv contained a C-terminal stop codon at residue 218 (or residue 163 of SHBs) and is designated $LHBs^{t218}$/ $MHBs^{t218}$ (clone 27.2) or at residue 227 (or residue 172 of SHBs) and is designated $LHBs^{t227}$/ $MHBs^{t227}$ (clone 3.1).

3.4.4. PHYLOGENETIC ANALYSIS

43 surface gene sequences were aligned with 49 sequences from GenBank and phylogenetic analysis was carried out (figure 3.25). All the isolates from HCC HBV infected patients belonged to subgenotype A1. Division into subgenotype A1 and A2 was evident. HBV isolates from HCC patients could not be separated phylogenetically from isolates derived from southern African asymptomatic carriers and acute hepatitis patients. The length of the surface gene of 43 subgenotype A1 isolates (24 full-length and 19 subgenomic isolates) varied between 984 bp and 1 203 bp.

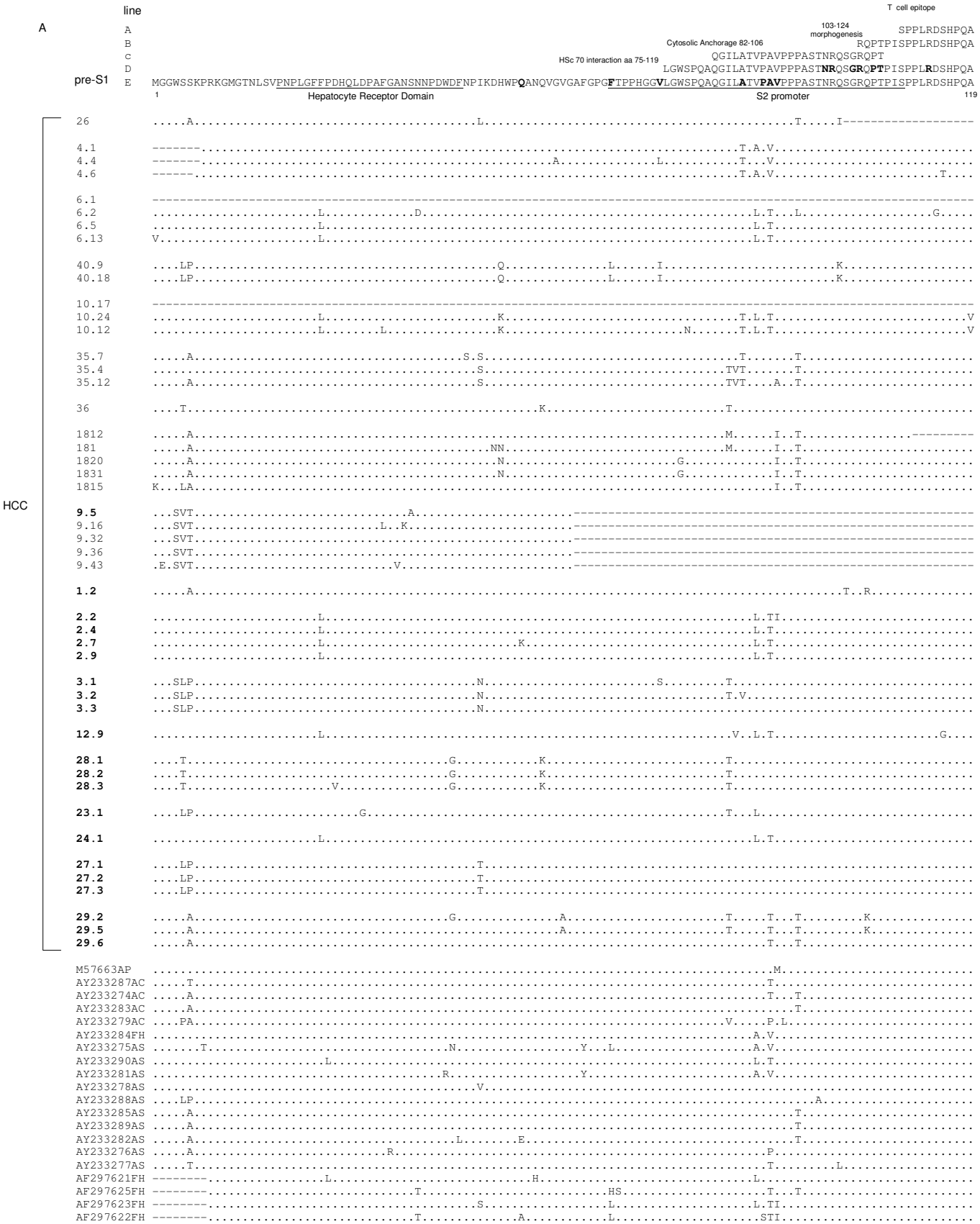

A
line
A
B
C
D
E
pre-S1
T cell epitope
103-124
morphogenesis
Cytosolic Anchorage 82-106
HSc 70 interaction aa 75-119
Hepatocyte Receptor Domain
S2 promoter
1
119
HCC
26
4.1
4.4
4.6
6.1
6.2
6.5
6.13
40.9
40.18
10.17
10.24
10.12
35.7
35.4
35.12
36
1812
181
1820
1831
1815
9.5
9.16
9.32
9.36
9.43
1.2
2.2
2.4
2.7
2.9
3.1
3.2
3.3
12.9
28.1
28.2
28.3
23.1
24.1
27.1
27.2
27.3
29.2
29.5
29.6
M57663AP
AY233287AC
AY233274AC
AY233283AC
AY233279AC
AY233284FH
AY233275AS
AY233290AS
AY233281AS
AY233278AS
AY233288AS
AY233285AS
AY233289AS
AY233282AS
AY233276AS
AY233277AS
AF297621FH
AF297625FH
AF297623FH
AF297622FH

Figure 3.22A

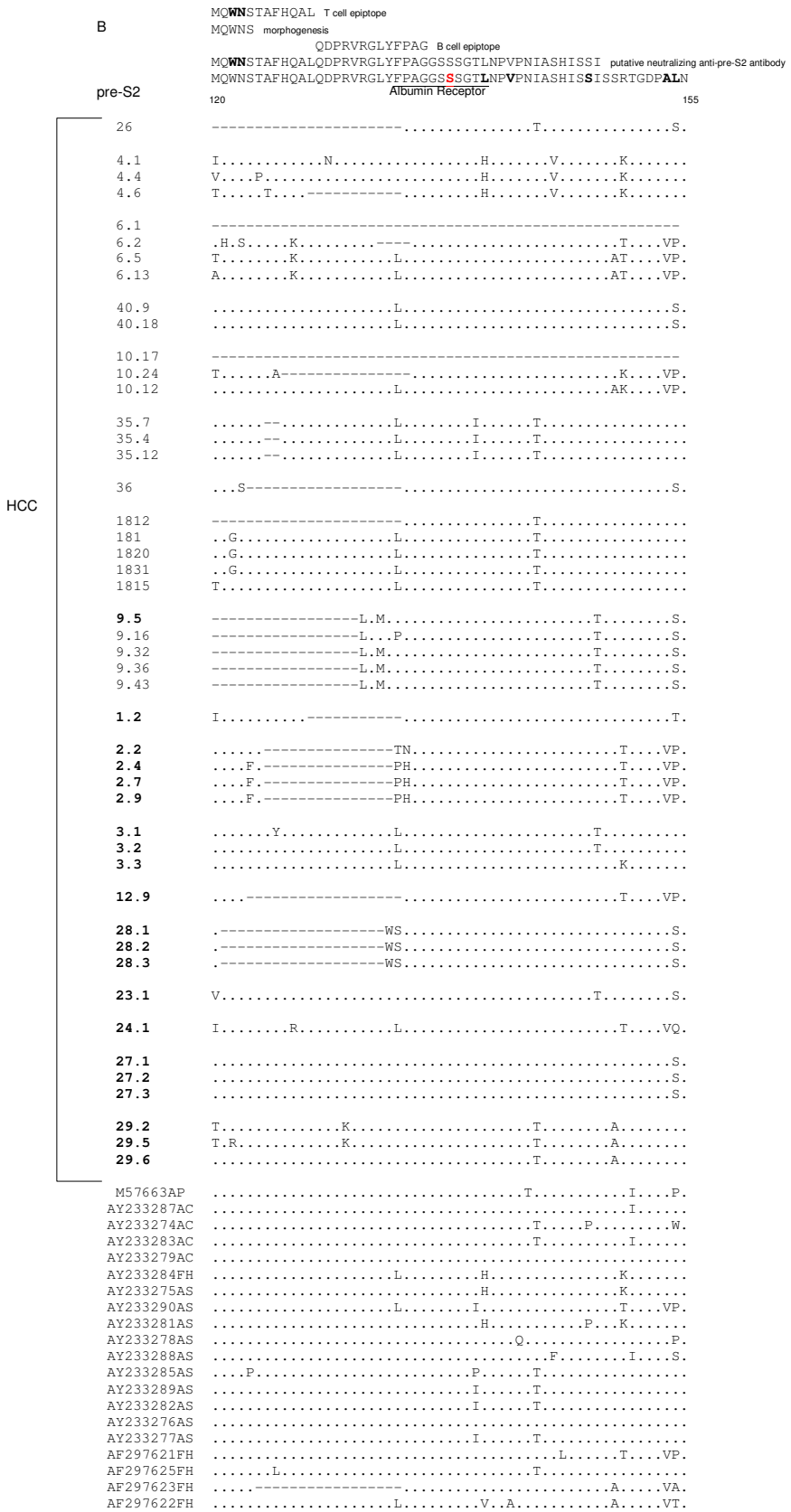

B
MQWNSTAFHQAL T cell epiptope
MQWNS morphogenesis
QDPRVRGLYFPAG B cell epiptope
MQWNSTAFHQALQDPRVRGLYFPAGGSSSGTLNPVPNIASHISSI putative neutralizing anti-pre-S2 antibody
MQWNSTAFHQALQDPRVRGLYFPAGGSSGTLNPVPNIASHISSISSRTGDPALN
pre-S2
120
Albumin Receptor
155
HCC
26
4.1
4.4
4.6
6.1
6.2
6.5
6.13
40.9
40.18
10.17
10.24
10.12
35.7
35.4
35.12
36
1812
181
1820
1831
1815
9.5
9.16
9.32
9.36
9.43
1.2
2.2
2.4
2.7
2.9
3.1
3.2
3.3
12.9
28.1
28.2
28.3
23.1
24.1
27.1
27.2
27.3
29.2
29.5
29.6
M57663AP
AY233287AC
AY233274AC
AY233283AC
AY233279AC
AY233284FH
AY233275AS
AY233290AS
AY233281AS
AY233278AS
AY233288AS
AY233285AS
AY233289AS
AY233282AS
AY233276AS
AY233277AS
AF297621FH
AF297625FH
AF297623FH
AF297622FH

Figure 3.22B

Figure 3.22 Amino acid sequence alignment of pre-S region: pre-S1 (A) and pre-S2 (B). The amino acid numbering below the consensus sequence is for LHBs or pre-S. Forty seven surface gene isolates from this study and 20 GenBank sequences are compared here. The consensus sequence of HBV subgenotype A1 is shown in line E. The consensus sequence for subgenotype A1 was generated using GenBank sequences listed within this figure (M5766AP-AF297623) and BioEdit with a 55% majority at each position, except at highly polymorphic positions where the most common amino acid was used. Lines A, B, C and D are used to illustrate the positions and sequences of important pre-S functional domains. These include Hepatocyte Receptor Binding domain (R), S2 promoter, Chaperone interaction domain (H), Cytosolic Anchorage Determinant (C) and morphogenesis domain (M). Immunogenic epitopes T, B and P are also included. The S (highlighted in red) corresponds to S28, critical in LHBs and MHBs transactivation activity. Subgenotype specific aa's are indicated in boldface (bf) in line E. Well conserved amino acids in the M domain are indicated in bf in line D. Sequences derived from full-genomes are indicated in plain text and those from subgenomic amplicons in bold face. Isolates from the same patient are group together and separated from other groups by a line spacing. Dots are used to indicate the same aa as the consensus sequence and dashes deleted sequences. Isolates from hepatocellular carcinoma (HCC) patients, asymptomatic chronic carriers (AS) patients, acute hepatitis (AC) and fulminant (FH) are included in the alignment.

Table 3.4.1. Deletions in pre-S domain and nucleotidet position

Pre-S1	Pre-S2	Pre-S1/S2	Pre-S1/pre-S2/S	Amino acids	Domain Affected	No. of clones	Clone
2855-2872	-	-	-	1-7	I	2	4.1 & 4.4
2852-2869	-	-	-	1-6, 131-141	I B	1	4.6
-	7-12	-	-	126-127	P	3	35.4, 35.7 & 35.12
-	48-56	-	-	139-142	B P	1	6.2
-	22-54	-	-	131-141	T B P	2	1.2 & 4.6
-	14-55	-	-	128-142	T B P	1	10.24
-	9-50	-	-	126-140	T B P	4	2.2, 2.4, 2.7 & 2.9
-	2-52	-	-	124-141	T B P M	1	12.9
-	3115-42	-	-	121-139	T B P M	3	28.1, 28.2 & 28.3
-	1-54	-	-	124-141	T B P M	1	36
-	-	3183-54	-	111-141	T B P M H	1	18.12
-	-	3153-54	-	102-141	T B P M H C	1	26
-	-	3040-37	-	61-137	(T B M H C) * P	5	9.5, 9.16, 9.32, 9.36 & 9.43
-	-	-	2064-486	1-155	(T B P M H C R)*	1	10.17
-	-	-	1935-871	1-155	(T B P M H C R)*	1	6.1

I - pre-S1 initiation start codon abolished

R - Hepatocyte Receptor Binding domain 21-47 (Neurath et al., 1986b)

H - Chaperone (Hsc70) interaction domain 75-119 / 63-107 (ayw) (Loffler-Mary et al., 1997)

C - Cytosolic Anchorage Determinant 82-106 / 70-94 (ayw) (Loffler-Mary et al., 1997)

M - morphogenesis domain 103-124 (Bruss, 1997)

T - T cell epitope 109-123 & aa 120-132 (Milich et al., 1987; Milich et al., 1986)

B - B cell epitope 133-145 (Milich et al., 1986)

P - putative neutralizing anti-pre-S2 antibody 120-145 completely deleted

Table 3.4.2 Mutations within the Virion Morphogenesis Domain (aa 103-124)

Amino Acid Δ	No. of Clones	Clone
Trp122Gly	3	18.1, 18.20 & 18.31
Trp122Arg	1	29.5.
Asn123Ser	2	6.2 & 36
Gln100Ile	1	26
Gln100Lys	2	40.9 & 40.18
Ser101Thr	1	1.2
Gln104Arg	1	1.2
Asp114Gly	1	6.2
Ser115Thr	1	4.6
Ala119Val	2	10.12 & 10.24
Met120Thr*	5	4.6, 6.5, 18.15, 29.2 & 29.5
Met120Ile*	3	1.2, 4.1, 24.1
Met120Val*	2	4.4, 23.1
Met120Ala*	1	6.13
Gln121His	1	6.2

* pre-S2 initiation codon abolished in 11 clones

Similar amino acids Ala=Val, Ile &Leu; Glu=Asp; Ser=Thr

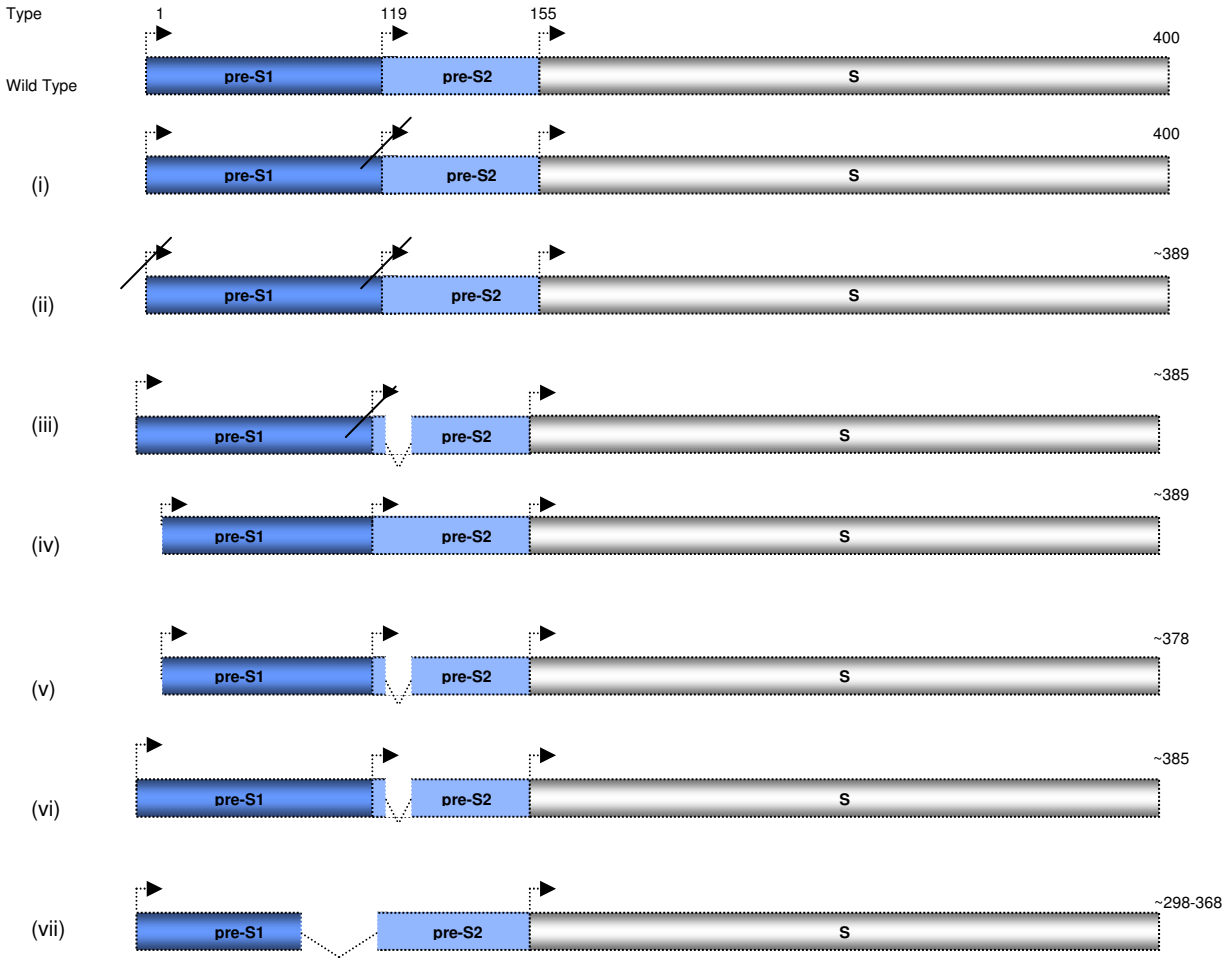


Figure 3.23 Schematic representation of pre-S mutational patterns found in the southern African subgenotype A1 HBV isolates. Amino acid (aa) positions are indicated above the wild type, the approximate size of LHBs is indicated on the right. The arrows indicate initiation sites, cancelled arrows indicate initiation codon mutations, broken lines represent deleted sequences. Seven types of deletion patterns were identified in this study.

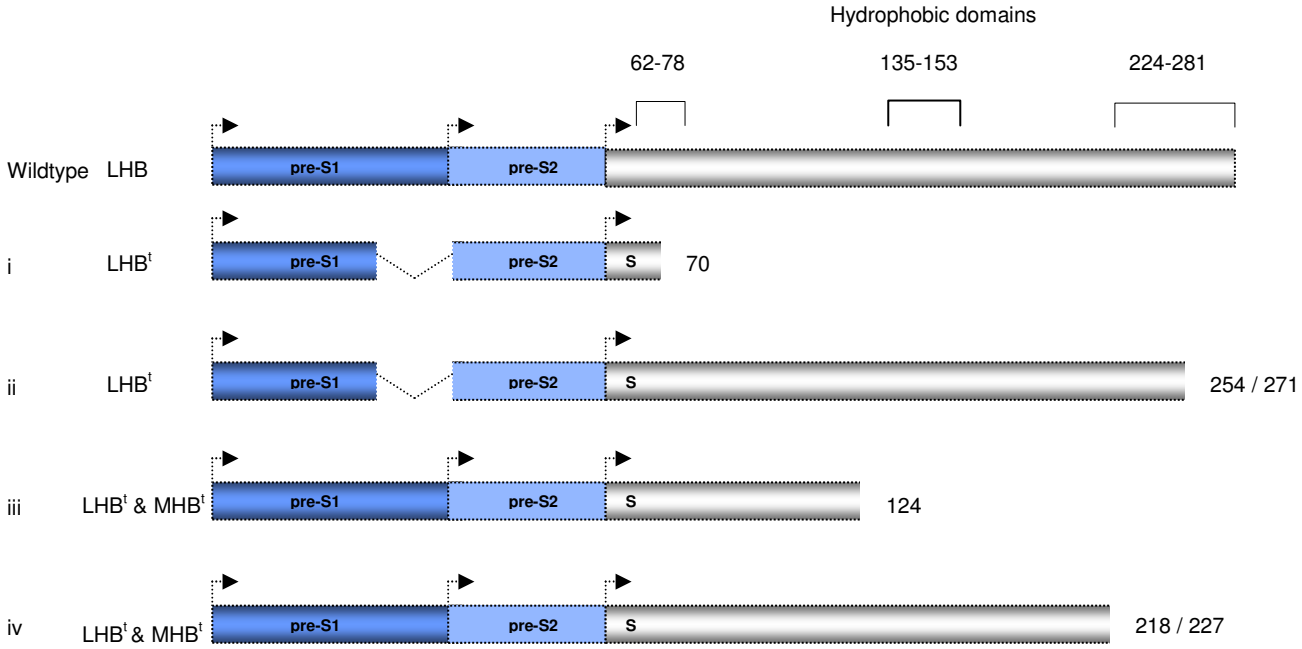


Figure 3.24 Schematic representation of LHBs, LHBs 3' truncated mutants (LHBs^t) and MHBs 3' truncated mutants (MHBs^t). Amino acid positions of MHBs are indicated in this figure. Arrows represent intact initiation codons. Broken lines represent deleted sequences and 3' truncations are indicated at the various amino acid positions. Four truncation patterns were identified in this data set. MHBs mutants would not be expressed from type i and ii due to the deletion of the pre-S2 initiation codon. Amino acids 224-281 are highly hydrophobic and deletion of these 70 aa is the minimum requirement for activation of pre-S2 transcriptional activation.

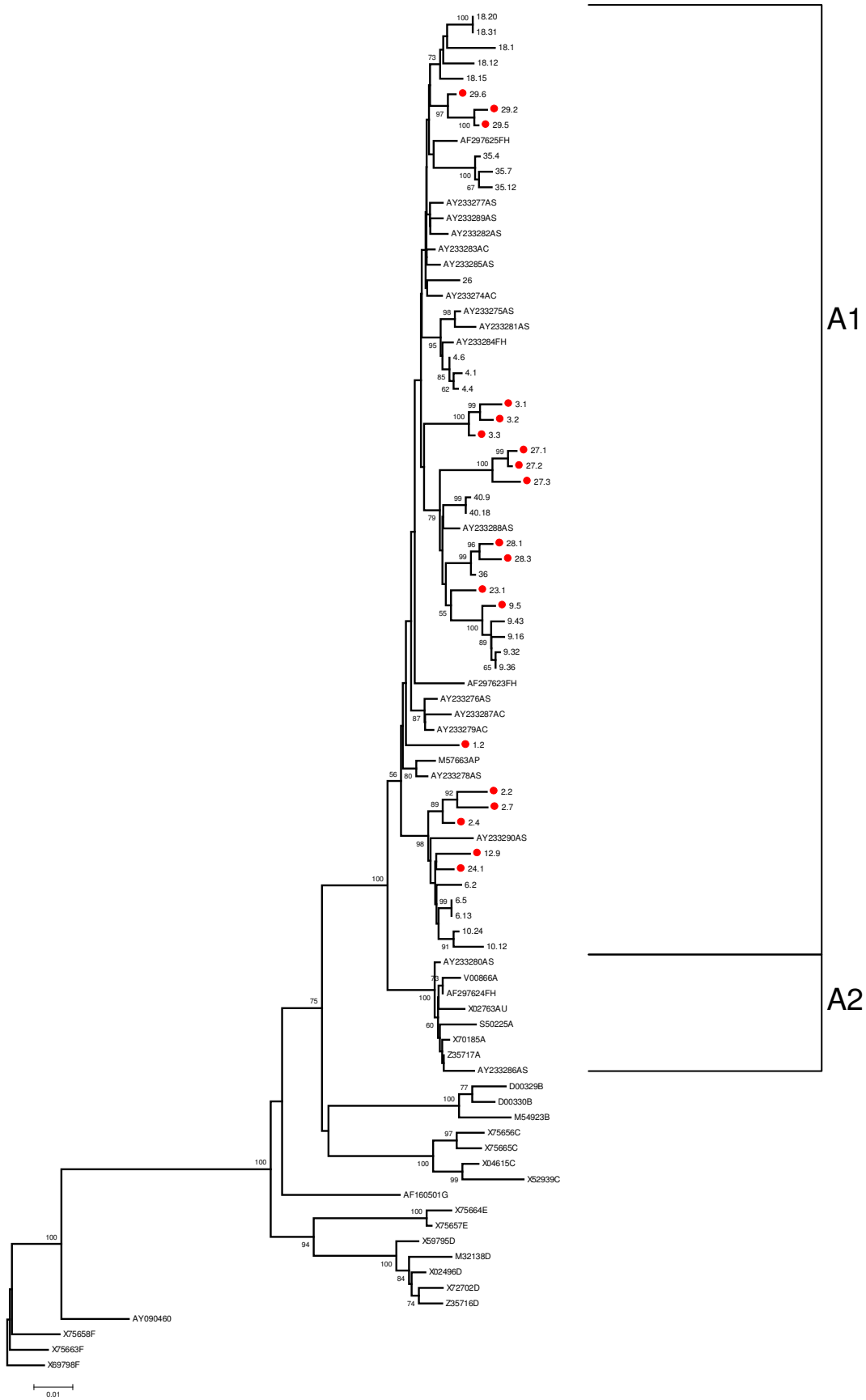


Figure 3.25. Surface gene dendogram including sequences from subgenomic isolates

3.5. FUNCTIONAL ANALYSIS OF MUTANT PRE-S/S GENES

Ground glass hepatocytes (GGHs) are a histopathological hallmark of chronic HBV infection and HCC, which are classified according to their integrated pre-S/S mutant genes (Fan et al., 2001; Hsieh et al., 2004a). In addition the majority of HCCs also contain integrated rearranged HBV pre-S/S genes (Kekule et al., 1990; Schluter et al., 1994) and pre-S mutants have been associated with progression to HCC (Huy et al., 2003; Sugauchi et al., 2003a). It has been established that cytosolic and membrane bound $MHBs^t$ and LHBs encoded by mutant pre-S/S genes form part of the pre-S2 family of transactivators that leads to the activation of PKC and the c-Raf1/Erk2 signal transduction pathway which increases cell proliferation (Kekule et al., 1990; Hildt et al., 1996a; Hildt et al., 2002). Additional evidence is accumulating, which implicates these variants in the disruption of several cellular growth and proliferation regulatory pathways, reviewed by (Wang et al., 2003).

In this study novel mutant pre-S/S genes were described in the episomal HBV isolated from the serum of southern African black HCC patients (section 3, figure 3.24). In a preliminary pilot study, selected pre-S/S genes characterised from this cohort of patients (figure 3.25) were cloned into pcDNA 3.1/His A/B/C expression vectors and transfected into HepG2 cells. P21 $^{Waf-1}$ protein levels and caspase 3 activity were measured as markers for cell cycle arrest and apoptosis, respectively.

The pre-S/S genes of HBV clones 1.2, 2.2, 9.5, 12.9 and 24.1 (figure 3.25) were subcloned into pcDNA 3.1/His A/B/C expression vectors and designated pcHBsAg1.2, pcHBsAg2.2, pcHBsAg9.5, pcHBsAg12.9 and pcHBsAg24.1

respectively. The pcHBsAg24.1 construct was used as a wild-type representative as it did not contain internal pre-S1/pre-S2 deletions or a 3' truncation introduced by a nonsense mutation (Δpre-S1/pre-S2LHBs^{t}). These clones were sequenced to confirm homology to the parent sequence and correct orientation. These constructs were transfected into HepG2 cells and p21^{Waf-1} was detected using western blot analysis and caspase 3 activity was measured using a fluorogenic substrate DEVD-AMC.

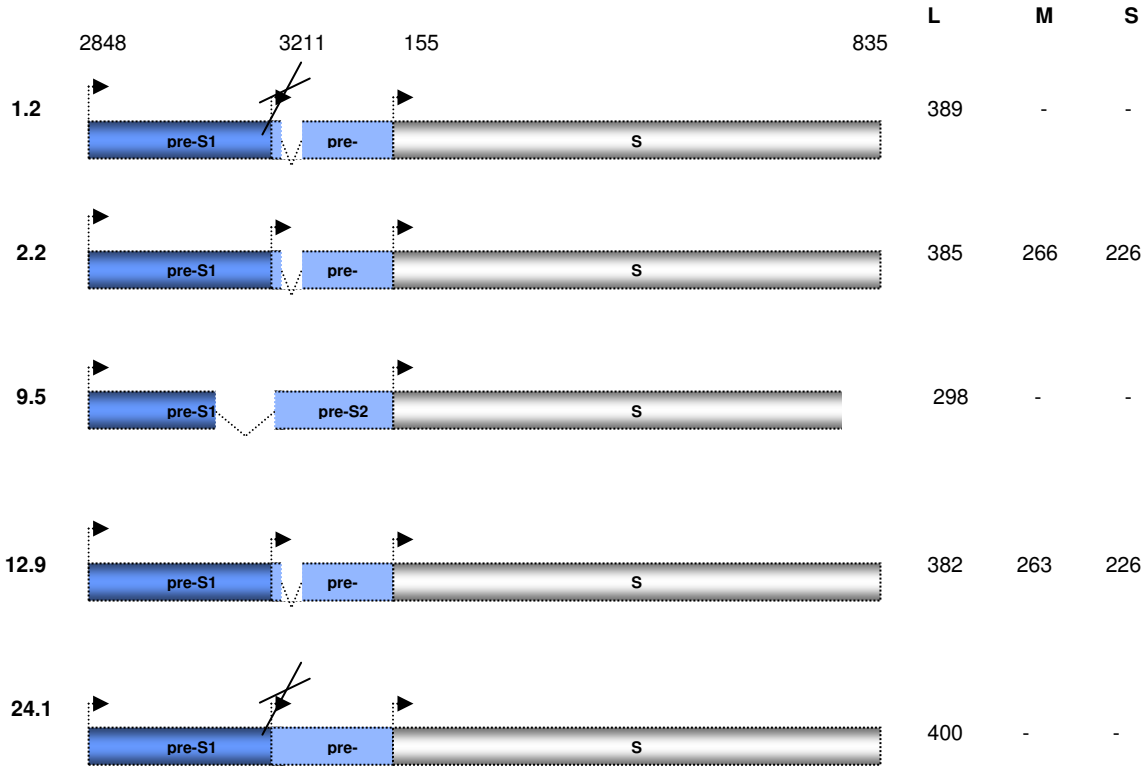
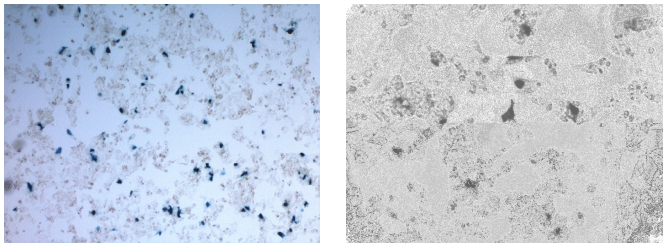
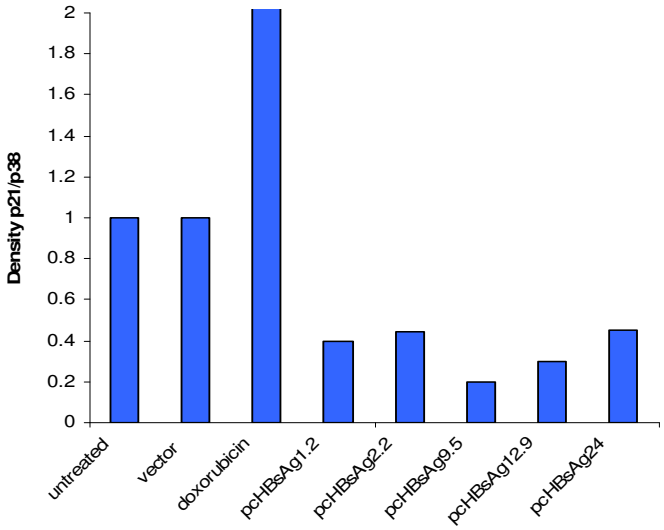


Figure 3.26 Schematic representation of the preS/S gene constructs (pre-S variants). The predicted amino acid (aa) lengths are given on the right hand side for the large (LHBs), middle (MHBs) and small (SHBs) surface proteins. No protein expression (-). Arrows indicate functional start codons, while cancelled arrows indicate abolition of a start codon.

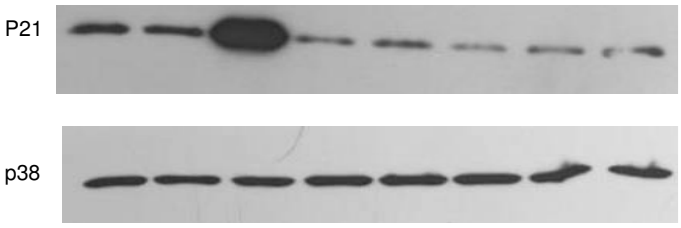
A



B



C



D

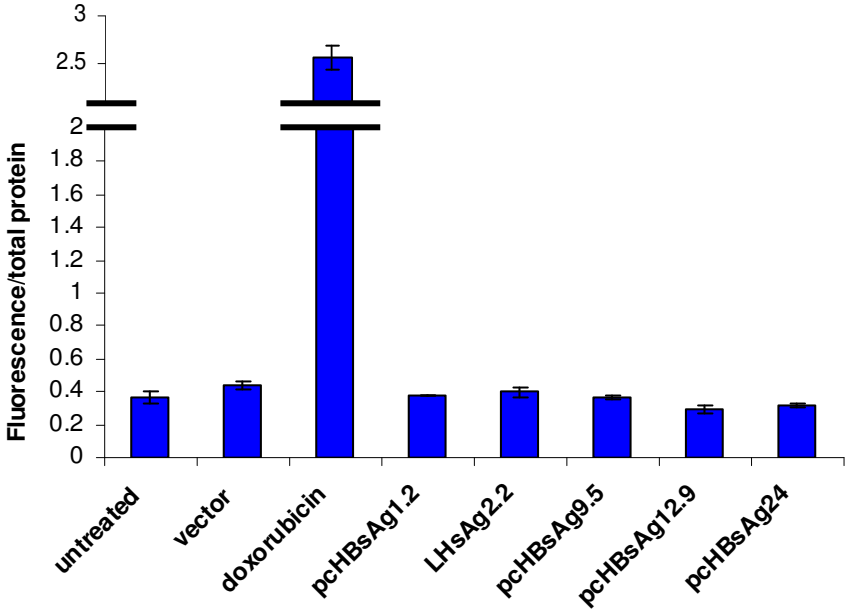


Figure 3.27

Figure 3.27 (A) Transfection efficiency was measured in parallel using X-gal staining, images of cells containing the indigo precipitate were captured with a Zeiss Axio Skop40 microscope, and showed an efficiency of 10-20%. (B) Western blot analysis of cell lysates from cells transiently transfected showed that $p21^{Waf-1}$ protein levels were decreased in HepG 2 cells transfected with all the pre-S/S mutant constructs. Lane1, untreated (untransfected); lane 2, Vector (pcDNA 3.1/His A); lane 3, cells 2 µM doxorubicin; and lanes 4-5, constructs pcHBsAg1.2, pcHBsAg2.2, pcHBsAg9.5, pcHBsAg12.9 and pcHBsAg24.1. 20 µg of cell lysate was examined for $p21^{Waf-1}$ protein levels by western blots using a $p21^{Waf-1}$ specific antibody (Santa Cruz Biotech). p38 was used as a protein loading control (C) Caspase 3 activity of cell extracts from transiently transfected cells was determined using a fluorogenic substrate aspartate-glutamate-valine-aspartate-7-amino-4-methyl coumarin (DEVD-AMC). Caspase 3 activity appeared to be only marginally affected by the pre-S/S mutants.

CHAPTER 4

4.0 DISCUSSION

4.1 SEQUENCE CHARACTERISTICS AND PHYLOGENETIC ANALYSIS OF HBV FROM BLACK SOUTHERN AFRICAN HEPATOCELLULAR CARCINOMA PATIENTS

4.1.1. HBV HETEROGENEITY

Complete HBV genome and sub-genomic surface gene amplification were carried out on DNA extracted from sera of 40 HBsAg-positive Black South African patients with HBV-induced HCC. When the complete genome and subgenomic PCR results were combined, 90% (36/40) of these patients were positive for HBV PCR. When evaluating the PCR fragment sizes, 58% (21/36) of the patients demonstrated fragment size heterogeneity of the complete genome amplicons (0.5 - 3.2 kb) and/or the surface gene amplicons (0.9 - 1.2 kb) (Table 3.1.1). PCR fragment size heterogeneity was evident both within and between patients. The presence of variants or quasispecies was confirmed within the serum of patients by cloning and sequencing. The size of the complete genome amplicons ranged from 1 120 bp to 3 253 bp, with the correct size of 3 221 bp for genotype A also being amplified.

HBV fragment size heterogeneity was present in relatively young patients: patient 39 (15 years old); patient 23 (24 years old) and patient 18 (28 years old) and in old patients (patient 10 of 75 years old). This suggests that variants can emerge over a relatively short period of time and persist over long periods of time, during end stage liver disease. Pre-S deletion mutants were present in both the young (patient 18) and old (patient 10) patients. It is clear that HBV

variants are present despite age differences and this observation can suggest a possible role for HBV variants in the development of HCC in younger patients. HBV-induced HCC is associated with black southern Africans of younger age, where genotype A infections prevail. A pre-S deletion variant has been identified in a child as young as 12 years with chronic persistent hepatitis B, although this variant did emerge a month after IFN-α treatment (Gerner et al., 1999; Gerner et al., 2003). A significant correlation has also been established between HBV pre-S deletion variants, older age and the development of HCC in genotype C (Sugauchi et al., 2003a).

4.1.2. GENETIC RELATESNESS AND PREDICTIVE MARKERS FOR HCC

Phylogenetic analyses was carried out for all the isolates characterized in this study, including complete genome isolates (section 3.1), surface gene amplicons (section 3.4) and two rare variants, isolates 6.1 and 10.17 (section 3.3), The analyses showed that all the HBV isolates characterized belonged to subgenotype A1. This is in complete agreement with an earlier study which implicated subgenotype A1 as the major genotype A strain involved HBV-induced HCC in black sub-Saharan Africans (Kew et al., 2005), despite the co-circulation of genotypes A, D and E in this geographical region. Similarly in Asia where genotypes B and C predominate, genotype C is associated with an increased risk of developing HCC (Yuen et al., 2004; Orito et al., 2005). Genotype B also exhibits regional differences in Asia with regards to disease association. In Taiwan HCC occurs more often in younger patients (Kao et al., 2000), whereas in Japan HCC occurs less often and in older patients (Orito et al., 2005). This may be because of the different subgenotypes, B1 (Japan) and B2 (mainland Asia), which predominate in these regions (Sugauchi et al., 2002).

Phylogenetic trees generated with the complete genome sequence, polymerase, pre-C/core, surface genes and X gene sequences (figures 3.1-3.5) all showed the HCC HBV isolates group within subgenotype A1. When using the complete genome, polymerase and surface gene sequences, the HCC HBV isolates did not separate from previously characterised subgenotype A1 isolates from ASC, AC and FH Black southern African patients (Owiredu et al., 2001b; Kimbi et al., 2004). The complete genome clones derived from individual HCC patients (4, 6, 9, 10, 18, 35, 40) formed minor monophyletic branches within the subgenotype A1 branch, which indicate that the clones were homologous. The position and distribution of the HCC HBV isolates within the subgenotype A1 branch was similar regardless of whether the complete genomes, the polymerase and surface genes were analyzed (figures 3.1, 3.2 & 3.4). In these dendograms, HBV isolates from patients 6 and 10 group together at the top of the branch, those from patients 40, 36 and 9 group together closely in the middle and from patients18, 35 and 26 at the bottom. The monophyletic branch of the HBV clones from patient 9 was more diverse than the isolates from other patients for the polymerase and surface genes. This is probably because the HBV genome from patient 9 contains a large deletion across the pre-S region (nt 3040 - 37) that would overlap with the polymerase gene and could explain the difference in position within both of these dendograms.

In the pre-C/core gene dendogram there appears to be a shift in position for HBV from patients 9, 36 and 40, which still group closely together but now at the upper end of the branch (figure 3.3). In addition isolates from patients 35 and 36 appear slightly removed from the rest of the isolates in the branch and

therefore perhaps are different to the rest of the isolates in this region of the genome.

In the X gene dendogram (figure 3.5) all the HCC isolates clustered together in the upper half of the subgenotype A1 branch together with AY233290, AY233275 and AY233281 (figure 3.5). These 3 isolates are from ASC individuals that are presumably clinically healthy. Interestingly, isolate AY233290 groups with isolates from HCC patients 6 and 10 in the complete genome and all the ORFs AY233275AS and AY233281AS group with isolates from patient 4 in all the ORFs and for the complete sequence (figures 3.1-3.5).

A closer look at the sequences of the X gene of these isolates (figure 3.7) revealed that all the HCC isolates and the 3 above mentioned ASC isolates have mutations in common. All have missense or deletion mutations involving either nucleotide 1762 and 1764 or both, within the C-terminal of the X gene, which overlaps the BCP. The 1762T1764A mutation was reported to occur more frequently in southern African black HCC patients (66%) than in ASC (11%), implicating this mutation in the development of hepatocarcinogenesis (Baptista et al., 1999). Several studies have reinforced the link between this double mutation and increased risk of developing HCC. In a prospective study, which followed 1638 high-risk individuals in Qidong (China) the 1762T1764A mutation emerged in 8 of 15 HCC patients (53.3%) before the cancer developed (Kuang et al., 2004). The 1762T1764A mutation was shown to be predictive of HCC in HBeAg-negative subgenotype C2 (Ce) infected Japanese individuals (Tanaka et al., 2006). Furthermore, in Chinese individuals the 1762T1764A mutation occurring in subgenotype B2 (Ba), was associated with HCC (Sung et

al., 2008). Interestingly, subgenotype B2 (Ba) is a genotype B/C recombinant in which the pre-C/Core region originates from genotype C (Sugauchi et al., 2002). A recent study based on the statistical analysis of 457 complete HBV genomes, also showed that the 1762T1764A mutation was associated with genotype C and G and not with genotype B and F (Kramvis et al., 2008). These studies all suggest that the 1762T1764A mutation could be a predictive marker for HCC in certain high risk populations infected with specific genotypes, underpinning the clinical significance of determining HBV genotype/subgenotype and mutation patterns. Therefore a longitudinal follow-up of these 3 South African ASC would be very useful to determine whether HCC does develop and would further support the current evidence for 1762T1764A as a predictive marker for HCC in southern African Blacks infected with subgenotype A1.

At a regulatory level the 1762T1764A mutation affects pgRNA and pre-C mRNA transcription by affecting transcription factor binding sites. The 1762T1764A mutation creates an imperfect HNF1 site (Li et al., 1999), specifically enhances pgRNA synthesis, and increases viral replication (Buckwold et al., 1996). These mutations have also been linked to suppressed HBeAg expression (Takahashi et al., 1995; Kurosaki et al., 1996; Baptista et al., 1999), through decreased binding of the LEF to the TA-enriched binding domain at nt position 1758-1764 (Buckwold et al., 1997). This mutation could contribute to the low HBeAg levels observed in subgenotype A1 infected individuals (Tanaka et al., 2004).

The HBeAg shares antigenic epitopes with the core protein and because it is expressed on the surface of hepatocytes and secreted into the blood, it may shield the core protein from a direct cellular and humoral immune attack. In the

presence of low serum HBeAg levels, the core protein may become the primary target. *In vivo* studies also show that the core protein elicits a more severe antibody response compared to HBeAg (Milich et al., 1988). These factors may induce necroinflammatory disease, a known contributing factor for hepatocarcinogenesis. If on the other hand anti-HBe seroconversion is achieved, this also leads to hepatocyte necrosis through the recognition and attack of hepatocyte membrane bound HBeAg. This severe anti-HBeAg immune pressure provides the setting for the emergence of HBeAg negative variants or variants with reduced capacity to express this antigen.

At the protein level the 1762 A→T leads to a Lys to Met change at residue 130 and the 1764 G→A leads to a Val to Ile change at residue 131. These residues occur close to or within the Kunitz domain-like sequence at the C-terminal end of HBx (aa 131-142) and may affect the transactivational capacity of HBx. Sirma *et al* (1999) examined naturally occurring HBx mutants derived from non-tumour and tumour tissue and showed that these possessed amino acid changes within the transactivation domain. Functional analysis showed a loss of transactivation activity and the ability to induce anti-proliferative pathways and instead a cell growth advantage was gained. One of these mutants also contained missense mutations Lys130Met and Val131Thr (Sirma et al., 1999). The study also demonstrated that mutations close to or within the either Kunitz domain-like sequences (aa 61-69 or 131-142) abrogated transactivation.

The global distribution of HBV genotypes and subgenotypes has been established, and it is becoming apparent that mutations associated with specific genotypes also have distinct global distribution pattern and may affect the

clinical outcome of infection within a specific populations (Kao et al., 2000; Kramvis and Kew, 2005).

According to statistical analysis carried out in a recent study there is a strong correlation between serological subtype *adw* and genotypes A, B, F, G and H, and *adr* with genotype C, and *ayw* with D and E (Kramvis et al., 2008). At a subgenotype level, serological subtype *adw*2 mainly occurs in subgenotype A1 and A2 and *ayw* in subgenotype A3 and A4. In the current study, the majority of patients (11/18) had HBV with serological subtype *adw*2, three had *ayw*1, and the remaining 3 had *ayw*2, even though they were all infected with subgenotype A1. This unexpected combination, subgenotype A1 and serological subtype *awy*1, has been identified previously in two subgenotype A1 isolates from South Africa (Kimbi et al., 2004) and two from the Philippines (Kramvis et al., 2008). *ayw*1 *is* also the serological subtype of subgenotypes A3 and A4, of West African origin.

These data suggest that serological subtype *ayw*1 may be more prevalent in subgenotype A1 in southern African than previously reported. Serological subtypes are not constrained by HBV genotype/subgenotype and hence do not have the same geographical distribution. In a similar unusual finding serological subtype *ayw* was identified in five genotype C isolates from Australian Aborigines, two with *ayw*3 and three with *ayw*1 (Sugauchi et al., 2001). These were mainly chronic hepatitis patients. However, one patient with *ayw*3 had liver cancer. It should be noted that unusual genotype and serological subtype combinations do occur in HBV isolates from HCC patients although their relevance is not clear.

When comparing HBV isolates derived from HCC patients to those derived from ASC, AC and FH patients, higher nucleotide divergences were found in both the complete genome and 5 regions of the genome in HCC-derived isolates (Table 3.1.2). The age range of the two groups of patients was similar: HCC patients (15 - 75 years) and ASC, AC and FH patients (25 - 56 years). This observation suggests that isolates present in HCC patients are under more pressure to evolve and avoid detection by the host immune system, which continually tries to eradicate the virus. The emergence of variants often coincides with flares in alanine aminotransferase (ALT) levels during long standing chronic infection (Santantonio et al., 1992; Nagasaka et al., 1998). The difference in nucleotide divergence was particularly pronounced in the pre-C/core and HBsAg and to a lesser degree in the pre-S region (Table 3.1.2). This is reasonable since the core, pre-S and HBsAg contain major immune epitopes and would be targeted by the cellular and humoral immune responses. Approximately 67% (6/9) of the patients have between 1 and 2 amino acid changes in the core protein B-cell epitope, which occurs between aa 70-90 (Salfeld et al., 1989). Eighty nine percent (8/9) of the patients have between 1-6 amino acid changes between residues 100-160 of HBsAg. Interestingly, patients 10 and 35 had one isolate each with Met133Thr mutation, which could possibly compromise antibody neutralization and may represent potential vaccine escape mutants (Carman et al., 1997). However, to our knowledge these individuals were not vaccinated for HBV and therefore the mutation described probably emerged as a result host immune pressure. This is further supported by the fact that residue 145Gly, a hot spot for vaccine escape, was conserved in all the patients. When combining the patients that had complete genome data and surface gene data, 94%

(17/18) of the patients had a HBV with either a deletion or amino acid substitution in the pre-S T and B-cell epitope region (section 3.4, figure 3.22).

4.1.3. SUBGENETYPE A1 CHARACTERISTICS AND SIGNATURE SEQUENCES

As established previously amino acids differentiating subgenotype A1 and A2 from each other and from other genotypes were concentrated in the pre-S1 region overlapping the spacer of the polymerase (figure 3.6) (Kimbi et al., 2004). There were sequence similarities in the pre-S1 region between isolates from HCC patients and those from ASCs, AC and FH patients.

Subgenotype A1 isolates have Gln54 in the pre-S1, which differs to subgenotype A2 with Ala54, genotypes B, C, E, F and G, which have charged acidic residues Asp or Glu and genotypes F and H, which have Met. Residue 54 occurs within the accessory domain of the pre-S1 involved in virion infectivity (aa 12-88) (figure 1.4). This residue is flanked by Pro on the amino terminal side and Ala on the carboxy terminal side (figure 3.22A). Pro is a hydrophobic amino acid with a cyclic structure usually found in the bends of folded protein chains and is usually accompanied by a negatively charged amino acid in the pre-S1 domain at position 54. It is not clear whether the loss of charge would affect protein function or structure. Residue 74 also occurs in the domain involved in virion infectivity subgenotype A1 and A2 both have hydrophobic residues at these positions, all the other genotypes have a basic Lys. This difference may also have an impact on infectivity. Residues 90 and 91 occur within the Hsc 70 interaction domain and the cytosolic anchorage domain (figure 3.22A). It is very curious that subgenotype A1 isolates appear to have either hydrophobic (Val) or

hydrophilic (Thr) amino acid at position 90 and essentially hydrophobic (Val) amino acid at position 91. These domains are involved in the dual topogenic capacity of LHBs, which facilitate both virion infectivity and nucleocapsid envelopment (Bruss, 2004).

Residues 32, 35 and 47 of pre-S2 occur within a putative neutralizing anti-pre-S2 antibody domain (figure 3.22B) and variation within this recognition site may promote immune escape. The subgenotype A1 specific amino acids in the S and polymerase priming region are well conserved among HBV isolates from HCC patients, ASCs, AC and FH patients.

In the polymerase spacer region (Thr^{236}, Gly^{268}, Tyr^{269}, Gln^{334} and Lys^{338}) there is consensus between the HCC isolates and the other subgenotype A1 isolates from ASCs, AC and FH patients listed in figure 3.6. Subgenotype A1 has specific differences at residues 334 and 338 (Kimbi et al., 2004) in the positively charged DNA binding cleft (Das et al., 2001) within the fingers of the HBV polymerase, which could affect viral replication. Subgenotype A1 has Gln^{334} and Lys^{338}, subgenotype A2 has Lys^{334} and Glu/Asp^{338} and Glu^{334} and Glu/Asp in all other genotypes. Glu and Asp are acidic, negatively charged residues. Subgenotype A1 has a neutral polar residue Gln^{334} and a positively charged basic residue Lys^{338} and this differs to the other genotypes listed. This may be associated with low HBV DNA levels characteristic of subgenotype A1. Interestingly, HBV from HCC 18 has Lys^{338} in one of five isolates and Gln^{338} in the remaining four (figure 3.6). HBV from HCC patient 35 has Gln^{338} in all its isolates (figure 3.6). In the majority of the isolates in these two patients, two neutral polar residues, Gln^{334} and Gln^{338} would “replace” charged residues in

the DNA binding cleft. It is not clear what effect this would have on DNA binding and ultimately replication.

The signature amino acid motif within the X region described previously in subgenotype A1 isolates was also identified in two HCC patients 'Ser^{11}, Ala^{31}, Ser^{47}, Ser^{146}, Ser^{147}' and is also present in the subgenotype A1 isolates from Malawi (figure 3.6) and the Philippines (Kimbi et al., 2004). The Ser^{146}, Ser^{147} within the X region result from mutations at nucleotide position 1809G→T and 1812C→T (figure 3.7). Three HCC patients had HBV with a variation of this motif 'Ser^{11}, Ala^{31}, Ser^{47}, Ala^{146}, Ser^{147}' and one had 'Ser^{11}, Ala^{31}, Ser^{47}, Ser^{146}, Pro^{147}. Whereas, patient 9 has 'Ser^{11}, Thr^{31}, Ser^{47}, Ala^{146}, Pro^{147}in all of its HBV isolates. Patient 9 maintains a hydrophilic amino acid at position 31 (Thr) similar to the other genotypes listed in figure 3.6 but quite different to subgenotype A2 with an Ala at this position. These changes do not occur within domains of the HBx protein, which are necessary functional transactivation activity (Arii et al., 1992).

Five of nine HCC patients and the subgenotype A1 isolate from the Philippines have C at position 1320 (figure 3.7). This differs to the other subgenotype A1 isolates but is similar to genotype D and H. It interesting that 1/3 isolates in patient 4, 2/3 isolates in patient 6, 5/5 isolates in patient 18, 1/1 isolate in patient 36 and 4/4 isolates in patient 9 had 1320C. It would appear that nucleotide A or C can be tolerated at this position in subgenotype A1. This nucleotide occurs approximately 50 bases upstream of the X gene initiation codon, within the 5' untranslated region of the X transcript and this variation may affect translational efficiency (Moolla et al., 2002).

4.1.4. MUTATIONS IN HBV PRE-C GENE AND BCP IN HCC IN SOUTHERN AFRICA

Although mutation at position 1862 was previously shown to occur in 29% of isolates from HCC patients (Kramvis et al., 1998) none of the HCC patients in the present study contained this mutation. This may be because only 9 HBV isolates were sequenced in the precore region in the present study.

The variation 1888G→A occurred in 3/9 (33%) HCC patients and 1888G→C occurred in 1/9 (~1%) which brings the total number of patients with a mutation at position 1888 to 4/9 (44%). This is similar to a previous study carried out in black South African patients with HBV-induced HCC using sub-genomic pre-C PCR analysis. The serum analysis revealed 1888G→A in 40%, 1888G→T in 8% and 1888G→C in 1% (Kramvis et al., 1998). 100% of tumours and 53% of non-tumorous tissue were also positive for a silent mutation at 1888 (Kramvis et al., 1998). Both 1888A/C are silent at amino acid level. However, 1888A stabilizes the encapsidation signal and *in vitro* analysis of this mutant shows that it does not disrupt viral replication and, does not affect HBeAg expression (Guarnieri et al., 2006). 1888G→A may however affect core protein expression as it introduces an out-of-frame AUG start codon 13 nt upstream of the core AUG start codon (Kimbi, 2005). This mutation is also unique to subgenotype A1 (Kramvis et al., 2008) and rarely occurs in other genotypes (Kramvis et al., 2008)

One of nine (~11%) of the HCC patients (patient 10) contained the 1896G→A mutation, which was always accompanied by the 1858C→T mutation in all clones including the splice variant. In agreement with previous findings (Kramvis

et al., 1997; Kramvis et al., 1998), the 1896G→A mutation is not prevalent in southern African patients infected with genotype A (subgenotype A1) because of the folding constraints of the encapsidation signal (Junker-Niepmann et al., 1990). Together 1858T1896A stabilizes ε and enhances viral replication but completely abrogates HBeAg expression (Carman et al., 1989). The 1858T and 1896A combination occurs naturally in genotypes B, D, E and C. This HBeAg negative mutant probably emerges due to host immune pressure and its emergence may coincide with HBeAg seroconversion.

In the present study, the BCP 1809G→T mutation was found in 43% of the HCC patient and the 1812C→T mutation was found in 57% and these mutations were found together in 29% of the HCC patients. In a previous study Baptista et al (1999) reported that 80% of South African isolates harbour similar mutations immediately upstream of the pre-C AUG codon. Double and triple mutations at nucleotide positions 1809-1812 are unique to subgenotype A1 and are stable genetic traits as they are found in HBV isolates from children and acute hepatitis patients (Ahn et al., 2003; Kimbi et al., 2004; Sugauchi et al., 2004a). These mutations may disrupt the optimal consensus Kozak sequence and may interfere with translation initiation of the HBeAg (Kozak, 1986). When these mutants were transfected in HepG2 cells, it was demonstrated the 1809T1812T moderately reduced HBeAg expression, whereas the triple mutants 1809T1811T1812T and 1809T1811C1812T drastically impaired HBeAg expression. Both of these triple mutants preferentially expressed p21, the core precursor and not p25, the HBeAg precursor. This could be attributed to leaky scanning of the precore AUG in these triple mutants. This is further supported by the findings of a case-control study, which found reduced HBeAg

and HBV DNA levels are characteristic of subgenotype A1 and that the former was linked to the Kozak sequence variations unique to this strain (Tanaka et al., 2004). Reduced HBeAg and increased core protein would presumably both contribute to enhance replication since HBeAg is known to inhibit viral replication (Lamberts et al., 1993; Guidotti et al., 1996b; Scaglioni et al., 1997) and core protein is essential for encapsidation and core particle formation. Analyses of sequences deposited in GenBank revealed that 1809T1812T is most prevalent in subgenotype A1 occurring in 76% of these strains (Kimbi, 2005; Kramvis et al., 2008).

In the present study, the 1762T1764A double mutation occurred in 78% (7/9) of isolates from HCC patients whereas 1764A alone occurred in one HBV from one patient. This is in agreement with the previous finding, which also demonstrated a high prevalence of 66% in black South African HCC patients (Baptista et al., 1999). Since, as discussed in section 4.1.2, the BCP mutations are prevalent in advanced liver disease and are often associated with the late HBeAg positive phase of disease or HBeAg seroconversion, investigators wanted to determine the virological impact of these mutations. Using a subgenotype A2 backbone and site-directed mutagenesis, various constructs were generated containing the BCP double mutants together with the less prevalent 1753C and 1766T mutations (Parekh et al., 2003) . Using transfection studies, the double mutant was shown to have a moderate effect on replication and HBeAg expression while the 1753C1762T1764A1766T quadruple mutant severely reduce HBeAg expression and greatly enhanced viral replication (Parekh et al., 2003; Tong, 2005). Similar findings were made in previous studies, which showed a gained replication advantage for the

1762A1764T variants compared to wild type from a chronic carrier (Kajiya et al., 2002) and an HCC patient (Lin et al., 2001).

The prevalence of pre-C and BCP in the complete HBV genomes from South African HCC patients is given in table 3.1.4. Five of nine (56%) patients had at least 3 of these mutations. In the four patients (patients 6, 9, 10 & 40) with more than one isolate, the mutations were present in all the clones, showing that these were the dominant mutations. It would appear that with the available knowledge about the effects of these mutations on viral replication and HBeAg expression that genomes isolated from patient 9, 10 and 36 would predictably have the highest replication capacity and greatly reduced HBeAg expression. However, overall there appears to be a tendency toward accumulating mutations which enhance viral replication and reduced HBeAg expression.

This finding is supported by a previous South African study, which demonstrated a high prevalence of similar pre-C and BCP mutation patterns in HCC patients using analysis of sub-genomic amplicons of the BCP region. The 1753C1762T1764A1809T1812T combination occurred in 10/50 (20%) HCCs vs 2/47 (4%) ASC. The 1753C1762T1764A1766T1768A1809T1812T occurred in 2/50 (4%) HCCs vs. 0/47 (0%) ASC. The 1764A1766T1768 A 1809T1812T occurred in 10/50 (20%) HCCs vs. 4/47 (8.5%) ASC (Baptista et al., 1999). The data indicate that the pre-C and BCP combination mutants may be more prevalent in HCC compared to ASC, however this should be investigated further. Furthermore, when pre-C and BCP mutations co-exist, HBeAg expression can be further reduced, suggesting that mutations, which affect

HBeAg translation and transcriptional regulation have a cumulative effect on the reduction of HBeAg protein synthesis (Ahn et al., 2003).

The 1766T and 1768A mutations have been associated with fulminant hepatitis, and *in vitro* analysis showed a 15 fold enhancement of virus replication by a posttranscriptional or translational mechanism involving elevated core protein production and enhanced encapsidation (Baumert et al., 1998). The effects of these mutations were shown within the context of an *adw* strain and were suggested to be strain specific (Baumert et al., 1998).

BCP mutations are often accompanied by mutations within other viral genes. Most of the isolates in the present study contained BCP mutations and pre-S deletions and missense mutations, which are known to cause intracellular surface protein accumulation and abrogate its secretion (Le Seyec et al., 1998; Gerner et al., 2003; Wang et al., 2003). Extensive variability in HBsAg expression and virion secretion amongst the 1762T1764A17531766T BCP variants has been observed and this could result in enhanced replication and concomitant reduced viral secretion within hepatocytes *in vivo*, which would contribute to the severity of liver disease (Kajiya et al., 2002; Parekh et al., 2003).

In this study using complete genome analysis, the 1753T mutation was present in HBV from 3/9 (33%) HCC patients and 1/10 (10%) ASCs. In an earlier study, using subgenomic amplification this mutation was found in isolates from 19/50 (38%) HCC patients and 6/47(13%) ASCs (Baptista et al., 1999). Statistical analyses showed 1753V(G/A/C) to be predictive for HCC in HBeAg-positive subgenotype C1 carriers (P = 0.0055), whereas 1653T was predictive for HCC

in HBeAg-positive subgenotype C2 carriers (P = 0.0118). In addition, the latter study also showed 1753V or 1653T or 1762T1764A was predictive of HCC in HBeAg negative subgenotype C2 carriers (P < 0.05). In the current study we also identified 1653T with the 1762T1764A in patients 26 and 35 (Appendix B figure B1). Perhaps these mutations should be investigated further in subgenotype A1 to determine whether these mutations could also be useful markers for early diagnosis of HBV-induced HCC in sub-Saharan Africa.

There appears to be concordance between the subgenotype A1 isolates from HCC patients and the other subgenotype A1 isolates listed in figure 3.7 when evaluating the subgenotype A1 specific changes in the S1 and S2 promoter regions. The subgenotype A1 isolates do have several unique variations within the *cis*-regulatory elements and these are listed in table 3.1.3. Nucleotide changes probably affect viral gene transcription and translation through these regulatory elements. In addition, due to the compact nature of the HBV genome nucleotide changes also affect the overlapping reading frames and effect amino acid changes within the viral proteins as discussed above. We can surmise that these mutations may contribute to the characteristically low HBV DNA levels found in subgenotype A1 infected patients, by modulating HBV replication. However, not all of these mutants have been studied *in vitro* and their effects remain to be determined.

4.1.5. CONCLUSION

The phylogenetic analysis of 24 HBV isolates from 9 HCC patients revealed that these were all of the subgenotype A1 strain. This is in complete agreement with previous findings in this region, which demonstrated despite the co-circulation of

genotype A, D and E in southern Africa that genotype A was mainly associated with the development of HCC (Kramvis et al., 1998) and that this could be attributed to subgenotype A1 (Kew et al., 2005). Interestingly phylogenetic analysis demonstrated three asymptomatic chronic carriers consistently clustering with the HCC isolates. This association was more pronounced with the X ORF and suggests that the BCP mutations common to all these isolates may be a useful predictive marker for HCC. This is supported by previous analysis which shows a significant association between the double BCP mutation and HCC in South Africa (Baptista et al., 1999).

Serological subtype *ayw*1 was also identified in three HCC patients and this is considered an unusual occurrence in genotype A (Kramvis et al., 2008). The combination between serological subtype *ayw*1 and subgenotype A1 has been identified previously in South Africa and Philippines and this may mean these strains are more prevalent than expected.

Detailed analysis of HCC isolates confirmed the presence of subgenotype A1 signature sequences, which were mainly concentrated in the pre-S region. HCC isolates did differ from the ASC, AC and FH isolates and these differences were concentrated in the BCP.

4.2 SIMULTANEOUS ALTERATIONS IN THE BASIC CORE PROMOTER, CORE GENE AND SURFACE GENE

4.2.1. HBV HETEROGENEITY WITHIN PATIENT 18

A complex BCP mutation was identified as the predominant mutation within the HBV genomes described in a young black African female of 28 years (patient 18) that developed HCC (figure 3.9). The BCP deletion disrupted the pre-C

promoter and together the deletion and insertion mutations introduced two HNF1 and one HNF3 transcription factor binding sites. In addition to the drastic rearrangements within the BCP, mutations were identified in the X, core and pre-S genes, that would result in the expression of dysfunctional truncated viral proteins (figures 3.12, 3.13B, 3.23 & 3.24).

4.2.2. PREVALENCE OF THE COMPLEX BCP MUTATIONS

All 5 HBV isolates derived from patient 18 contained the complex BCP mutation, whereas 4/5 isolates contained the single nucleotide deletion at nt position 2090 , 2/5 the X gene initiation codon mutation and only one isolate contained the 2 bp deletion and 45 bp insertion in the core gene and the pre-S gene deletion. The heterogeneity of the mutations found in the core, pre-S and X gene combined with the homogeneous BCP confirms that these mutations are not PCR artefacts. The predominance of the BCP mutation implies that the selection of this trait is beneficial for the survival of the virus. This finding does not exclude the possibility that wild-type virus is also present in this patient.

A gradual take over by the complex BCP mutants within mixed virus populations of chronically infected HBV patients has been demonstrated in several longitudinal studies (Preikschat et al., 1999; Fujiwara et al., 2005). BCP deletions and complex insertions emerge prior to the onset of severe liver disease and become the predominant virus population during cirrhosis and end stage liver disease (ESLD), and at this stage the HBV BCP mutants also accumulate additional mutations in the core and pre-S genes. In fact, a statistically significant correlation has been established between a combination

of BCP and core gene mutations with or without pre-S gene deletions and the progression to cirrhosis and ESLD (Preikschat et al., 2002).

This type of BCP and core gene mutation has not been previously described in black South African HBV infected patients. Sub-genomic screening of the BCP and pre-C/Core region in serum associated virus in large numbers of HCC (>50) and ASC (>40) patients did not reveal similar mutations (Kramvis et al., 1998; Baptista et al., 1999; Kimbi et al., 2004). Complete genome analysis of ASC and AC also failed to identify mutations of such complexity (Kimbi et al., 2004). However, similar mutations were found pre-C/Core sub-genomic region between nt (1814-1901) in tumour and non-tumour tissue of HCC patients (Kramvis et al., 1998). These included deletions varying between 27-72 bp and an insertion of 38 bp which was only detected in non-tumorous tissue. More than 50% of these mutations were also associated with topoisomerase I cleavage sites. These mutations invariably disrupted DR1 and ε and hence were not considered replication competent viruses, instead must have been amplified from HBV integrants. Although there were similarities between the previous findings and the current findings, these viruses were found in different compartments: in tissue in the previous study compared to serum in the current study, and in different regions of the genome pre-C/Core vs. BCP, respectively.

Although this type of mutant has not been described in South African HBV infected patients previously, it has been reported in numerous studies in HBV related conditions including asymptomatic and symptomatic chronic carriers (Barlet et al., 1994; Horikita et al., 1994; Laskus et al., 1994b; Okamoto et al., 1994; Moriyama et al., 1994), FH patients, immunosuppressed patients and in

patients undergoing IFN treatment (Fujiwara et al., 2005), from other geographical regions. Studies also show a significant association of these complex mutations with advance liver disease including liver cirrhosis and ESLD. Similarly to a previous study in a Korean HCC patient (Kim et al., 2007), we demonstrated that this mutant is the predominant mutation in an HCC patient, thus , providing further support for the role of this type of HBV mutant in advanced liver disease.

4.2.3. HOMOLOGOUS RECOMBINATION MEDIATED BY TOPOISOMERASE I

The BCP mutation described in all five clones from patient 18 resembled the complex rearrangements in clonal HBV DNA integrations found in HCCs (Wang et al., 2004). HBV integrants from human HCC usually contain a combination of inversions, deletions, direct duplications or inverted duplications. These HBV integrants can vary in size from 300 bp to 3 500 bp, much smaller and larger than wild-type HBV. It appears that the DR1, DR2 and cohesive overlap region cohesive overlap region (COR) are intricately involved in the generation of these integrants and mutants since the majority of virus-virus junctions in HBV mutants or virus-cellular junctions and virus-virus junctions in HBV integrants occur within or near these regions.

The molecular structure of the HBV genome and its replication pathway makes it susceptible to mutagenesis (Wang, 1985; Tuttleman et al., 1986a; Tuttleman et al., 1986b). HBV Dane particle DNA exists as a partially double stranded molecule with the COR being held together by hydrogen bonding, the DR1 in the DNA minus strand contains a nick and the region 5' to DR2 in the plus strand is single stranded. This partially double stranded DNA molecule is only converted into a covalently closed circular DNA molecule once it enters the nucleus. Before this conversion is achieved, the partially double stranded DNA molecule could succumb to the activity of topoisomerase I activity within the nucleus, because the DR1, DR2 and COR contain preferred topoisomerase I cleavage sites and these are the preferred site for virus integration (Dejean et al., 1984; Hino et al., 1986; Hino et al., 1986; Yaginuma et al., 1987; Hino et al., 1989).

Topoisomerase I is a cellular enzyme primarily involved in DNA replication and relieves the torsional stress of supercoiled DNA by the simultaneous cleavage and religation reactions on single or double stranded DNA (Wang, 1985). Topoisomerase I cleavage of partially double stranded DNA or double stranded DNA containing nicks in the non-cleaved strand or single stranded DNA substrates results in the uncoupling of cleavage and ligation reactions. DNA fragments with cohesive ends with topoisomerase I still attached via the 3' phosphate would facilitate recombination with available heterologous acceptor DNA strands (Wang, 1985).

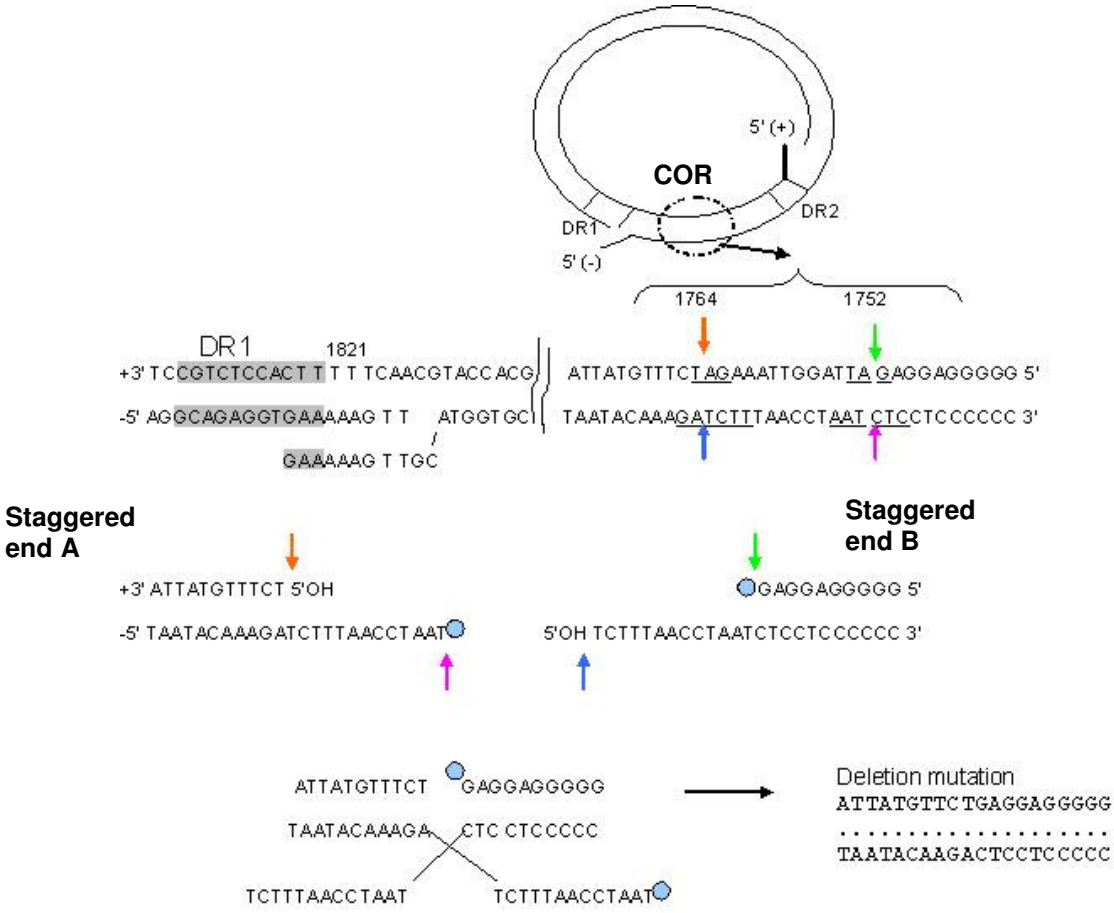


Figure 4.1. Proposed model of Topoisomerase I mediated recombination in the HBV strains isolated from patient 18. Adapted from (Wang and Rogler, 1991; Pourquier et al., 1999)

Figure 4.1 Proposed model of Topoisomerase I mediated recombination in the HBV strains isolated from patient 18. HBV virion DNA illustrated above, followed by the sequence and region of interest in the COR. The arrows indicate the potential topoisomerase I cleavage sites with preferred recognition sequences. After cleavage, staggered ends A and B are produced. The opposite end of the plus strand contains a free 5' OH group (end A) which becomes linked to 3' bound topoisomerase I at staggered end B. The excess sequences are digested by cellular nucleases to produce the final deletion mutation. The dots represent bound topoisomerase I.

Based on the topoisomerase I mediated WHV DNA integration *in vitro* model (Wang and Rogler, 1991) and the DHBV model (Pourquier et al., 1999), the involvement of topoisomerase I mediated HBV mutagenesis/homologous recombination could explain the mutations detected in the HBV genomes isolated from patient 18. Topoisomerase I appears to be directly involved in the creation of the 13 bp interstitial deletion since the left and right cleavage sites at nt 1752 and nt 1764, respectively, contain preferred topoisomerase I motifs (Konopka, 1988). Topoisomerase I cleavage intermediates could be created by the cleaving of opposite HBV DNA strands within the COR of the same molecule (figure 4.1). This would require at least two cleavage reactions per cohesive end. In the DNA cohesive ends labelled B, cleavage may occur at nt position 1764 in the plus strand and nt position 1752 in the minus strand and similarly in the opposite strand. In the DNA cohesive ends labelled C, cleavage may occur at nt position 1752 in the plus strand and at nt position 1764 in the minus strand. The number of base pairs between the DRs and the respective cohesive ends should be sufficient to maintain base pairing to hold the DNA molecule together with the cohesive ends available to function as donor and acceptor ends in an intra-strand reaction. In this model cleavage at nt 1752 in the plus strand would free the 5' OH and act as the acceptor end allowing it to link to the 3' bound topoisomerase at nt 1764. The single stranded overhanging fragments would presumably be digested by cellular nucleases and the ligation reaction completed by topoisomerase I. This would then result in the observed deletion between nt 1752 and 1764 (figure 4.1).

In the WHV DNA integration model, WHV DNA cleaved in the minus strand opposite the nick, caused strand linearization and provided the acceptor strand. The cellular DNA fragment with cohesive ends and covalently bound

topoisomerase I provided the donor strand, and together these successfully illustrated heterologous recombination (Wang and Rogler, 1991). The *in vitro* DHBV model also illustrated plus and minus strand topoisomerase cleavage reactions. DHBV minus strand topoisomerase cleavage and intra-strand religation reactions also demonstrated the introduction of 1 and 2 bp micro-deletion mutations and minor duplications (Pourquier et al., 1999).

The more complex mutations such as the insertion at nt 1770 could also be mediated by topoisomerase I (figure 3.10). This is supported by the fact that the internal duplications are enriched with A/T-rich regions and these are also topoisomerase I preferential recombination sites (Konopka, 1988). In addition a CTT motif occurs in the middle of the two larger duplications and homologous sequences flank the entire insert (5'-GGAGG-3') (figure 3.11B). So it would appear that several complex rearrangement events occur in the creation of this mutation pattern. Similarly the core gene micro-deletions (nt 2078, 2079 and 2090) and internal duplications appear to involve several topoisomerase I cleavage and religation reactions (figure 3.13A).

The data show that the virus-virus junctions described in the BCP and the core gene of the HBV genomes from patient 18 occur in the vicinity of topoisomerase I recognition motifs. These results are consistent with the findings for WHV, where virus-virus junctions of complex rearrangements found in four WHV genomes isolated from serum of a chronically infected woodchuck were closely associated with topoisomerase I recognition motifs (Kew et al., 1993). In addition, the overwhelming majority (~92%) of hepadnaviral recombination sites from integrants

found in HCC DNA contained topoisomerase I recognition sites (Schirmacher et al., 1995).

Since these complex rearrangements identified in virion associated genomes are similar to those found in integrated HBV DNA from HCCs it is feasible to suggest that HBV genome mutagenesis could occur prior to HBV integration and not exclusively after integration. Infection with HBV characteristically results in a rapid turnover of hepatocytes due immune attack, cell necrosis and regeneration and this is accompanied by increased cellular DNA replication. Since topoisomerase I is primarily involved in replication, it would be available during this phase at higher concentrations to facilitate DNA replication and inadvertently cause HBV mutagenesis and integration into the host genome (DiNardo et al., 1984).

4.2.4. BCP REARRANGEMENTS

The pre-C promoter was most affected by the BCP insertion and deletion mutation. The initiation sites for pre-C mRNA (nt 1785 -1786 and nt 1791-1797) and the pgRNA transcription (nt 1815 ± 5) remained intact (figure 3.10 A & B). The TATA-box like motifs TA1, TA2 and TA3 underwent considerable rearrangement (nt 1760 to 1780) (figure 3.10 A & B), even though all three motifs are needed for pre-C mRNA transcription. TA2 is situated at the optimal downstream position ~25-30 bp from the pre-C initiation site TBP binding. The deletion completely abolished TA1 and TA2 and the insertion created a TA2 duplication and 3 TA3 repeats. The TATA-box like motifs were rearranged so that the first TA2 and TA3 motifs are juxtaposed a mere 9 bp downstream from the pre-C initiation site (figure 3.10B), this could affect the binding of the TBP and its ability to initiate the assembly of the transcription machinery including the binding and positioning of

RNA polymerase II, which is needed for transcription. In support of this, the majority of BCP deletion mutation described previously, disrupt the TA2 and TA3 motifs (Okamoto et al., 1994) and *in vitro* analysis using primer extension assays demonstrated the abolition of transcription initiation at the pre-C initiation site and a 5-30% decrease in pre-C transcript production with a concomitant reduction HBeAg expression (Gunther et al., 1996b).

Remarkably, the TA4 motif remained intact in all five clones from this HCC patient. This motif serves as the pre-C mRNA initiation sequence, as well as the pgRNA TATA - box like motif (CATAAATT) (Yu and Mertz, 1996). This motif is also conserved in all the complete genomes (24 complete genomes) described in all the HCC patients in this study. This underpins the indispensability of this binding site as it is critical for pgRNA transcription initiation and hence HBV replication. Several studies, which described similar BCP mutants consistently show that the TA4 motif remains very well conserved (Okamoto et al., 1994; Gunther et al., 1996b).

The deletion at 1752-1764 disrupted Sp1, HNF4, TBP and LEF binding sites and the insertion mutation displaced the remaining sequences further downstream from their original position (figure 3.11A). Together the deletion, insertion and substitution (C1766G) (Table 3.2.1) mutations have introduced a potential HNF3 site and two HNF 1 sites (figure 3.11B). The putative HNF1 GGTTAAATATTAGGA motif has been described previously (Laskus et al 1994 a & b, Repp et al1992, Gunther et al 1996) as specifically being introduced through the deletion of nucleotides 1763 - 1770. In comparison our deletion occurred at nucleotides 1752 - 1764 and did not appear to be directly responsible for generating the HNF1 sites.

When comparing three HNF1 sites from various regions within the HBV genome there does appear to be sequence homology (figure 3.11C). Affinity assays have shown that the BCP HNF1 site described here has an affinity for HNF1, which is slightly lower than that of pre-S1 HNF1 motif (Gunther et al., 1996b).

The introduction of duplicated HNF1 sites and one HNF 3 site may have a drastic effect on viral replication and gene expression. Similar mutants demonstrate a 250-750% increase pgRNA/C mRNA and an associated increase in released HBV DNA containing viral particles following *in vitro* analysis (Gunther et al., 1996b). These mutants also show decreased HBeAg and envelope protein expression and increased core and polymerase protein expression. The observation that all the isolates (5/5) in patient 18 contained the same BCP mutant and the fact that previous evidence indicates that these mutations would confer a growth advantage compared to the wild-type, suggests that these mutants have replaced the wild type as the dominant quasispecies, by virtue of the gained replication capacity.

In addition to the increased core protein (HBcAg) levels expressed by HNF1 mutants, nuclear and cytoplasmic core protein accumulation has also been demonstrated in the liver of a fulminant hepatitis patient (Pult et al., 1997), these findings were further supported by independent *in vivo* and *in vitro* analyses of the HNF1 mutation (Fujiwara et al., 2005). Hepatocytes “overexpressing” HBcAg would be particularly susceptible to specific cytolysis because HBcAg elicits a strong cytotoxic T cell response (Penna et al., 1991) and in this way contributes to disease progression. Increased viral polymerase protein levels could also contribute to severity of disease because it too elicits a potent cytotoxic T lymphocyte response (Rehermann et al., 1995).

4.2.5. THE CORE GENE AND PRODUCTS

In addition to the potential altered levels of protein expression caused by the BCP mutation, which is common to all five genomes described in patient 18, 4/5 genomes also contain a mutated core gene compared to 1/5 with a wild-type core gene (genome 18.15) (figure 3.9). Genome 18.12 also contained a mutated pre-S gene with an in-frame deletion across the pre-S1/pre-S2 region as well as a premature stop codon at residue 15 of the S gene (figure 3.9, 3.23 & 3.24). Similarly, 18.1 contained an S gene premature stop codon at residue 69. The core protein expressed from genome 18.12 would be truncated after Asn 64 and in genomes 18.1, 18.20 and 18.31 after Asn 63 (figure 3.13B). These truncated proteins have the entire C-terminal arginine-rich domain deleted, which is involved in DNA packaging and contain less than half of the first 147 aa involved in nucleocapsid formation. The core protein expressed from genome 18.15 would be of wild-type length (185 aa) (figure 3.13B).

Truncated core proteins may interact with wild-type core proteins to form mosaic/recombinant viral particles. These particles may or may not be able to interact with viral DNA. Carboxyl-end truncated core proteins fused in-frame to the C-terminus of HBsAg drastically inhibited viral replication by interfering with virion assembly (Scaglioni et al., 1994). The proposed antiviral activity of truncated core proteins may lead to the accumulation and intracellular retention of viral proteins and may eventually contribute to cytotoxicity. Interestingly, several studies have identified core mutants with in-frame internal deletions (Gunther et al., 1996a; Preikschat et al., 2002; Fujiwara et al., 2005). At the correct ratio, genomes containing internal core gene deletions can improve the virulence of wild-type HBV (Gunther et al., 2000).

4.2.6. THE PRE-S/S GENE AND PRODUCTS

The mutations occurring in the pre-S and S genes of genomes 18.1 and 18.12 would result in an array of truncated envelope proteins. Carboxyl-end truncated LHBs, MHBs and SHBs would be expressed from genome 18.1, whereas only truncated LHBs and SHBs would be expressed from 18.12 since the pre S1/pre S2 internal deletion abolishes the pre-S2 initiation codon. The possible transactivators expressed would be LHBs^{t124} and MHBs^{t124} generated by an S gene stop codon 69 from isolate 18.1 (figures 3.9 & 3.24) and LHBs^{t70} generated by an S gene stop 15 from isolate 18.12. (figures 3.9 & 3.24)

The effects of these mutated proteins are two fold. Firstly, pre-S1 deletions interfere with transcriptional regulation of envelope expression, this dysregulation precipitates into the intracellular accumulation and retention of envelope proteins (Xu and Yen, 1996; Bock et al., 1997; Melegari et al., 1997). The retention of mutated and probably misfolded proteins has also been shown to activate the protein UPR pathway (Wang et al., 2003). Furthermore, in transgenic mice, HBV envelope protein retention results in cytotoxicity as the hepatocytes became sensitive to cytokine-mediated injury and further immune attack (Gilles et al., 1992; Nakamoto et al., 1997).

MHBs^{t124} have been classed into the pre-S2 family of transactivators, whereas LHBs^{t124} and LHBs^{t70} remain functionally uncharacterised. Similar transactivators have been described previously: the membrane bound MHBs^{t76} and MHBs^{t167} (Lauer et al., 1992; Hildt et al., 1996b) have a cytoplasmic orientation in the ER membrane to facilitate transactivation, whereas MHBs^{t52}, which contain no hydrophobic transmembrane domains has a cytosolic localization (Hildt et al.,

2002). The association between these transactivators and PKC results in the phosphorylation of Ser28 within the pre-S2 region and the concomitant activation PKC. This leads to the activation of c-Raf-1/MEK/ERK (extracellular signal regulated kinase) signal transduction pathway, which have been shown to increase hepatocyte proliferation (Hildt et al., 2002). In addition, LHBs and $MHBs^{t76}$ transgenic mice have been shown to develop HCC (Chisari et al., 1989). Furthermore, these C-terminally truncated pre-S/S genes have been described in integrated HBV in several HCC studies (Nagaya et al., 1987; Caselmann et al., 1990; Kekule et al., 1990) and one study identified these mutants in 25% of the HCC cases (Schluter et al., 1994).

4.2.7. THE X-GENE AND PRODUCTS

Interestingly all the isolates from patient 18 contained mutated X genes. All the genomes (5/5) contained a 3'-truncated X gene caused by the BCP deletion and insertion mutation, which resulted in a premature stop codon, leading to the truncation after Tyr125 (figure 3.12). In addition, isolates 18.20 and 18.31 contained X gene initiation codon mutations, completely abolishing HBx expression from these genomes.

Conflicting evidence exists for a role for wild-type HBx in hepatocarcinogenesis. Certain studies demonstrate the oncogenic effect of HBx in transgenic mouse models, while similar *in vivo* studies could not show HCC induction in other transgenic mouse lineages. Wild-type HBx can induce a late G1 cell cycle block in Chang cells, which was similar to the effect of tumour suppressor p53 (Sirma et al., 1999). Additional anti-proliferative characteristics include the inhibition of clonal outgrowth and induction of apoptosis in a p53 independent fashion. The same

study also showed that a naturally occurring HCC associated 3' truncated HBx protein, with 26 aa deleted at the COOH-terminal end, could abrogate the anti-proliferative effects of HBx by enhancing clonal expansion. This could be explained by the fact that the COOH-terminal end houses a growth suppressor effect domain (includes the Kunitz-type serine protease inhibitor domain) and upon deletion of this domain the oncogenic capacity of *myc* and *ras* is enhanced (Tu et al., 2001). Interestingly, approximately 80% of HCC tumours have been shown to contain rearranged X genes (including 3' deletions) caused by HBV integration (Wang et al., 2004). It is quite apparent from the current study and many others that X-gene rearrangements can be caused prior to HBV integration.

The 3' X gene deletions described in the isolates from patient 18 would result in a 29 aa deletion at the COOH-terminal end (figure 3.12) and by comparison to previous studies this type of mutant results in the loss of the anti-proliferative properties of the protein (Sirma et al., 1999). Furthermore, two isolates would not express HBx at all and in a sense this could be considered as the loss of a tumour suppressor gene. This is reminiscent of a similar finding in a genotype C strain isolated from a Korean HCC patient, where the entire X gene was deleted (Kim et al., 2007)

4.2.8. CUMULATIVE EFFECT OF MULTIPLE MUTATIONS IN ALL ORFS

The subgenotype A1 HBV genome population described in patient 18 presents a highly complex combination of mutations not previously described in HBV isolated from southern Africans. Within a single genome not only are the replicative regulatory elements greatly altered, to improve viral replication and reduce HBeAg expression, but deletions occur in every gene (core, surface, polymerase and X)

creating a myriad of truncated dysfunctional proteins which may have tumour promoting properties. A similar discovery has been described in a 48 year old Korean man with HBV-related HCC infected with subgenotype C2 (Kim et al., 2007). Four genomes were described in this patient, 2 of wild-type length and 2 with complex mutations involving similar BCP mutations but also deletions in all the viral genes, as well as an S gene premature stop codon. In addition all the genomes possessed mutations, which abolished HBeAg expression and consequently the patient was HBeAg negative. Unfortunately, the HBeAg status of the 28 year female in the current study was not available, however we do know that subgenotype A1 infected carriers have characteristically reduced HBeAg levels. Interestingly, a West African male infected with genotype E also developed an HNF 1 duplication in the BCP, this mutation was accompanied by a core deletion but no pre-S gene deletion, no 1896 mutation and the patient was HBeAg positive (Fujiwara et al., 2005). Fortunately, the patient did not advance to more severe liver disease due to a course of IFN monotherapy followed by lamivudine and interleukin 12. However, it remains unclear which additional mutations would have developed in the absence of treatment.

4.2.9. CONCLUSION

The description of these complex BCP mutations in the HBV strains in a 28 year old black South African female infected within subgenotype A1 is a novel discovery. These data also confirm that these types of mutations are not restricted to a specific genotype because they have been described in genotypes A, C and E.

The BCP promoter mutation described in the genomes isolated from patient 18 would clearly abrogate HBeAg expression and enhance viral replication. The loss of HBeAg expression leads to the elimination of the immune decoy and the redirection of immune attack to the HBcAg, expressed on liver cells, which may also have severe consequences.

When considering data from previous longitudinal studies and from the current study, it appears as the BCP mutation is the initial mutation, followed by deletions in the core gene then the pre-S gene (Preikschat et al., 1999). Additionally the current study identified S gene stop codon mutations, which have been documented previously (Kim et al., 2007). The combined effect of these mutations within one genome or within one virus population may contribute to more severe liver disease such as HCC.

The presence of 3'-terminally truncated pre-S/S genes, which also contained internal deletions within the pre-S region was demonstrated. These genes would not only produce $MHBs^t$ but $LHBs^t$ as well. The functional significance of the latter class of proteins has not been explored.

4.3. IDENTIFICATION OF RARE HBV VARIANTS IN HEPATOCELLULAR CARCINOMA

4.3.1. HBV HETEROGENEITY IN PATIENT 6 AND PATIENT 10

This section describes the detailed molecular structure of isolates from patient 6 and 10. Following complete HBV genome amplification the HBV isolates from both patients demonstrated a population of heterogeneous genome size (figure 3.15). Cloning and complete sequence analysis revealed the molecular structure of at least three types of genomes in each patient. Patient 6 contained a complex

combination of wild-type sequences, a pre-S deletion variant and a poly (dA) variant (figure 3.17) and patient 10 also contained wild-type sequences, a pre-S deletion variant and a splice variant (figure 3.19).

These results demonstrate the usefulness of complete genome amplification. Firstly, heterogeneous virus populations can be identified without direct cloning of virion associated HBV DNA, which requires high viral loads. Secondly, this method guarantees the identification of combinations of mutations on single genomes, which are more representative of *in vivo* sequences. Subgenomic PCR is very useful for identifying mutations within the target region, but reconstruction of complete genomes using sequences of subgenomic amplicons is labour intensive and may not be representative of *in vivo* viruses.

4.3.2. THE POLY (dA) VARIANT

4.3.2.1. GENOME STRUCTURE AND REPLICATION STRATEGY

The poly (dA) variant characterized in patient 6 (variant 6.1) is similar to the variants previously described consisting of internal poly (dA) tails ranging in size (60 - 100 nucleotides) with adjacent large internal deletions of varying length (1.8-2.9kb) (Sommer et al., 1997). The poly (dA) tail frequently starts at position 1935 just downstream of the authentic polyadenylation signal. Interestingly the 3'-end of the deletion in variant 6.1 occurred at nucleotide position 871, which is unusual because deletions in this type of variant have previously only been described at positions 612, 794, 966, 1622 and 1702 (Sommer et al., 1997) from the *EcoR*I site (Galibert et al., 1979).

The genome structure of these genomes implies that they are generated through partial DNA minus strand synthesis coupled with the reverse transcription of a poly

(dA) tail of an RNA pregenome. Normal reverse transcription begins with the viral polymerase/reverse transcriptase (RT) attached to the 5' ε structure of the pgRNA, a primer is synthesized and the polymerase/RT and attached primer translocates to the 3' ε structure where the primer anneals to complementary sequences within DR1. DNA minus strand synthesis begins with the simultaneous degradation of the pgRNA by the RNase H RT activity. The RNase H of the viral polymerase/RT degrades only that part of the pgRNA molecule, which is reverse transcribed and present in a DNA/RNA hybrid form (Sommer et al., 1997). The 3' end of the pgRNA, which is not degraded, also contains a DR1 sequence and poly (dA) tail. In the replication defective model, DNA minus strand synthesis is "paused" and the viral polymerase/RT switches templates and begins to reverse transcribe the poly (dA) tail of the non-degraded pgRNA fragment and proceeds to the 3' end of this fragment (Sommer et al., 1997). The length of the deletion depends on when during the minus strand synthesis the template switch occurs. Mutations in the RNase H catalytic domain result in pausing sites during DNA minus strand synthesis and these may correlate to the above "template switch" sites (Chen and Marion, 1996). Determining the DNA form of this genome has been hindered by its aberrant expression and low concentration, so it is not clear whether this molecule is single, partially double stranded or circular (Sommer et al., 1997).

4.3.2.2 RECOMBINANT PROTEINS

In the poly (dA) variant the deletion disrupts every HBV ORF except the X gene. In addition to the possible HBx protein expression from these genomes, a core-polymerase recombinant protein could possibly be expressed from a mutant pgRNA. Using GeneDoc computational analysis, a predicted protein of 280 amino acids containing an N-terminal core protein homologous domain (aa 1-11)

followed by a polylysine ‘tail’ (18 aa) and remaining protein sequence homologous to aa 599-844 of the viral polymerase RNase H domain was generated (figure 3.18). It is not known whether RNase H domain would be functional in this predicted recombinant protein. It is conceivable that gene expression does occur from these genomes because they have previously been detected in liver tissue and serum of infected patients (Sommer et al., 1997). It is not known whether virions containing these defective genomes are infectious. Although they contain all the regulatory sequences required for replication such as the BCP, encapsidation signal, DR1 and DR2 they would require wild-type surface, core and polymerase proteins.

In this study, a poly (dA) variant has been identified in 1 of 9 HCC patients, which is lower compared to the original study (Sommer et al., 1997), which detected these variants in 4 of 10 patients. It has also been suggested that poly (dA) variants emerge regardless of the host immune status because these variants have been identified in immunosuppressed, immune stimulated (IFN-α treatment) and fulminant hepatitis patients (Sommer et al., 1997). In the present study, we report the poly (dA) variant in a HCC patient. Poly (dA) variants may be uncommon in HCC. However, the detection of any variant can be hampered by the fact that the viral population can vary during the course of an infection and this could be caused by host immune pressure as well as the relatively short half-life of HBV in the serum of 1-3 days (Nowak et al., 1996; Zeuzem et al., 1997).

4.3.3. HBV SPLICE VARIANT

4.3.3.1. GENOME STRUCTURE

A HBV splice variant containing a 1 642 bp deletion, with deletion boundaries at nucleotide position 2067 and 489 has been characterized in a 75 year old male with chronic HBV infection who developed HCC (patient 10). Splice variants are found more frequently in the serum and liver tissue of chronic HBV carriers (Terre et al., 1991; Gunther et al., 1997a) than in the serum of patients with acute HBV infections (Rosmorduc et al., 1995). The presence of a virion associated spliced genome in patient 10 demonstrates that HBV pgRNA splicing occurs in the liver during advanced HBV-induced disease and that these genomes may be involved in hepatocyte transformation (Bonvin et al., 2006).

Splice variants represent a viral sub-population generated by the reverse transcription and encapsidation of spliced pgRNAs (Terre et al., 1991; Gunther et al., 1997a). A comprehensive analysis of splice variants was carried out using the complete genome amplification method and at least eleven types were identified (Gunther et al., 1997a). HBV splice variants are generated through the use of conserved splice donor and acceptor sites. Acceptor sites occur at nucleotide positions 2067, 2087, 2236, 2447, 2471 and 2985. Donor sites occur at nucleotide positions 282, 489, 2350 and 2902. The most common splice variant has a single intron removed at sites 2447/489. A maximum of three introns has been reported (2067/2350, 2447/2902 and 2985/489). Genotype A, D and E are unable to utilize 5’ acceptor site 2087 since these strains have a G residue instead of a T or C at the invariable +2 position of the splice site (Gunther et al., 1997a).

All splice variants described to date, including 2067/489 described in patient 10, lack the pre-S/S gene promoter sequences and adjoining coding regions for the polymerase and surface proteins but retain essential sequences required for viral replication and pgRNA/pre-C mRNA synthesis. Variant 2067/489 also has most of the core gene deleted and predictably defective pre-C and core proteins could be expressed from the resulting spliced RNA transcripts.

4.3.3.2. RECOMBINANT PROTEINS

The translated amino acid sequence of a core-polymerase fusion protein was predicted using the mutant core-polymerase gene sequence of splice variant 2067/489 (isolate 10.17) (section 3.3.7) which may be expressed *in vivo* from the spliced pre-C mRNA/pgRNA. The predicted protein was 433 aa long (figure 3.20). The first 56 aa were homologous to the N-terminal domain of the core protein and the rest was homologous to aa 469-844 of the viral polymerase, consisting of a partial RT domain and a complete RNase H domain. This protein may have RNase H activity. However the viability and function of such recombinant protein is unknown. Similarly pre-C fusion proteins could also be expressed from splice pre-C mRNAs. Recombinant fusion proteins have been shown to interfere with virion morphogenesis (Scaglioni et al., 1994). *In vivo* studies also show a higher ratio of splice variants to wild-type virus in the liver tissue compared to the serum of HBV infected individuals and WHV-infected woodchucks (Terre et al., 1991; Hantz et al., 1992). *In vitro* analysis revealed the intracellular accumulation of splice genomes (Sommer et al., 2000). This could lead to an abundance of defective pre-C/Core proteins within the cells. These proteins have the potential to abrogate viral replication and viral pregenome packaging (Scaglioni et al., 1994; Scaglioni et al., 1996; von Weizsacker F. et al., 1996; von Weizsacker F. et al., 1999). This

could eventually lead to the accumulation of all viral proteins and cause cellular stress.

Several studies have explored recombinant HBV proteins. Soussan et al described a 10 kDa HBV-spliced generated protein (HBSP) which is transcribed from a recombinant ORF (Soussan et al., 2000). HBSP starts with the polymerase initiation codon and the first 46 aa of the polymerase protein fused to 47aa corresponding to a new sequence generated by a frameshift mutation. *In vivo*, HBSP was detected in HBV-infected liver tissue and antibodies to HBSP were found in 30% of chronic carriers. More recently, HBSP epitopes were shown to elicit a multispecific T-cell response when tested with peripheral blood mononuclear cells obtained from chronic HBV carriers (Mancini-Bourgine et al., 2007). The expression of HBSP *in vitro* had no effect on viral replication and transcription, but was able to induce apoptosis. *In vitro* analysis of the splice variant 2447/489, revealed the expression of a 43 kDa polymerase-surface fusion protein in HuH-7 cells (Huang et al., 2000). This fusion protein could replace the LHBs in virion morphogenesis (Huang et al., 2000).

Another control point may exist which could regulate the flow of spliced mRNAs from the nucleus to the cytoplasm. All HBV RNAs also contain a post-transcriptional regulatory element (PRE), which is functionally similar to the HIV rev response element (Huang and Liang, 1993). PRE appears to regulate the ratio of spliced to non-spliced HBV RNAs and may be involved in the export of non-spliced RNA from the nucleus to the cytoplasm (Huang and Yen, 1995). The relatively high nuclear and intracellular concentration of spliced genomes (Sommer et al., 2000) may facilitate viral DNA integration into the host genome.

4.3.3.3. HBV SPLICE VARIANTS AND HCC

The human hepatoma cell line PLC/PRF/5 established from a tumour obtained from a black African patient with HBV-induced HCC (Macnab et al., 1976), contains an integrated HBV DNA with a deletion with nucleotide boundaries 2448-2901 (Ziemer et al., 1985). These boundaries coincide with splice acceptor/donor sites 2447/2902 (Terre et al., 1991; Gunther et al., 1997a). This suggests that infection with splice variant 2447/2902 led to viral DNA integration. In this study we identified variant 2067/489 in genotype A, interestingly variant 2067/489 has previously been shown to occur with variant 2447/2902 in a genotype A infected individual (Gunther et al., 1997a). So even though HBV splice RNAs have been shown to be dispensable for replication *in vitro* (Wu et al., 1991), there appears to be evidence for an indirect role in viral pathogenesis.

4.3.4. CONCLUSION

Complete genome amplification of serum derived HBV has revealed the presence of an unusual poly (dA) variant in patient 6 and a splice variant in patient 10. To our knowledge this is the first study to identify this type of virion associated variant in HCC and the first in black southern African patients infected with subgenotype A1. The role of the poly (dA) variant in HBV-related disease is not clear. We know that this defective genome is produced in hepatocytes by an unconventional replication strategy and encapsidated and secreted into the serum. The possible expression of HBx and defective viral proteins could contribute to pathogenesis. Patient 10 was one of the oldest cancer patients in this study, and it would appear that chronic infection with a heterogeneous population of viral genomes, including a splice variant have contributed to hepatocarcinogenesis.

Splice variants do not appear to be essential for virus life cycle, but may indirectly be involved in disease through the enhanced expression and intracellular accumulation of viral proteins. In addition there appears to be evidence of integration of spliced genomes and it may be useful to use known splice donor and acceptor sites to explore HBV integration patterns as integrants are notoriously difficult to characterize.

4.4. CHARACTERIZATION OF PRE-S/S GENE MUTANTS IN HCC IN BLACK SOUTHERN AFRICANS

4.4.1. PREVALENCE AND DISEASE

In the current study 72% of the HCC patients contained a deletion mutation in the pre-S region (pre-S1 and/or pre-S2), whereas none of the asymptomatic chronic carriers or acute hepatitis patients displayed this mutation. One patient with fulminant hepatitis did possess a deletion in the pre-S2 region (figure 3.22B). Recently, it has become apparent that these variants are associated with end stage liver diseases, such as cirrhosis and HCC (Takahashi et al., 1998; Gunther et al., 1999; Tai et al., 2002; Gunther, 2006). A significantly higher prevalence of these mutations in genotype C infected HCC patients compared to ASC has also been reported (Sugauchi et al., 2003a), with no difference in the prevalence of these mutants between genotype B and C (Huy et al., 2003). In an earlier study, none of the ASC (0/12) had these mutations compared to 50% of patients with CH (8/16), 50% (2/4) of the patients with cirrhosis, and 64% (7/11) of the patients with HCC (Ohno et al., 1993). It is clear that both pre-S1 and pre-S2 mutations accumulate during the late phases of persistent infection, however Gunther suggests that pre-S1 deletion mutations specifically emerge during active liver inflammation of HBeAg-positive patients and pre-S2 deletion mutations emerge during ESLD (Gunther et al., 1999; Gunther, 2006).

4.4.2. THE PRE-S1 AND PRE-S2 REGIONS

In this study pre-S1 deletions and initiation codon mutations were less prevalent compared to the pre-S2 deletions and initiation codon mutations (25% and 4% vs. 53% and 25%, respectively). The reason for this difference may be because of the different functions of the pre-S1 and pre-S2 domains. The N-terminal domain of the pre-S1 contains the hepatocyte receptor domain (aa 21-47). which is critical for attachment (Neurath et al., 1986b) and infectivity (De et al., 2001). This domain remains well conserved and uninterrupted by deletions in all the subgenotype A1 isolates listed in figure 3.22A except isolates 6.1 and 10.17, which have the entire S ORF deleted. Residues 42-47 have also been shown to be important for pre-S stability and maintaining the appropriate tertiary structure of pre-S domain (Lian et al., 2008). Deletions disrupting the S2 promoter are also uncommon, occurring in only 4/18 patients or 8/47 HCC-derived HBV isolates (6.1, 9.5, 9.16, 9.32, 9.36, 9.43, 10.17 & 26) (figure 3.22A). This mutation leads to the transcriptional dysregulation of envelope gene expression, which reduces the levels of MHBs and SHBs, negatively affecting virion assembly, eventually leading to the intracellular retention of envelope proteins (Gerner et al., 2003).

The pre-S1 domains appear much more susceptible to deletion mutations in genotype B and C (Huy et al., 2003; Sugauchi et al., 2003a). Sugauchi et al showed that deletions in the hepatocyte receptor domain and the S2 promoter occurred in approximately 50% and 90% of the patients infected with genotype B strains from diverse clinical backgrounds (ASC, CH, cirrhosis and HCC) and similarly in patients infected with genotype C (20% and 45%) (Sugauchi et al., 2003a). This could suggest that the pre-S1 gene in subgenotype A1 may be less

prone to mutation and could have a greater infectivity capacity during advanced liver disease compared to genotype B and C strains. However, no deletions in these domains in genotype C strains isolated from 40 HCC patients were found in another study (Takahashi et al., 1998). This implies that pre-S1 mutations are not as common in HCC and this is supported by the finding that pre-S1 mutations emerge during the relatively early stages of disease and may even be replaced by other dominant mutations during disease progression.

More than 60% (12/18) of the HCC patients in this study contain HBV strains with a deletion in the pre-S2 in at least one isolate. These deletions either occur across the pre-S1/pre-S2 boundary or within the first 5 residues of the pre-S2 N-terminal domain or a pre-S2 initiation codon mutation with or without a deletion occurring after residue 6. The pre-S1/pre-S2 boundary deletions would affect the critical domain involved in virion morphogenesis aa 102-124 (section 3.4 figure 3.22 A & B), ultimately affecting virion secretion (Le Seyec et al., 1998; Lian et al., 2008). The first five residues of the N-terminal domain further improve the stability of the pre-S protein domain (Lian et al., 2008) and this implies that they are involved in maintaining the correct secondary structure of the protein. Deletions in this domain would presumably result in misfolded proteins, which would also interfere with virion assembly.

Over 40% of the HCC patients contained HBV strains with pre-S2 initiation codon mutations. Amino acid changes at this position could affect the secondary structure of the pre-S and eventually change the stability of the protein. As shown in table 3.4.2, the initiation codon often changes from Met→Thr/Ile/Val/Ala. Thr is a

hydrophilic amino acid whereas IIe and Val are hydrophobic, which is similar to Met. Asn-4 (N), which is N-glycosylated in pre-S2 domain (figure 1.4 and 1.5) is essential for the role of MHBs in virus infectivity (Mehta et al., 1997) is replaced by Ser (S) in two patients specifically in isolates 6.2 and 36 (figure 3.22B). Other conserved amino acids within the virion morphogenesis domain are also substituted by amino acids with very different properties. Acidic amino acids such as Gln (Q) is replaced by hydrophobic amino acids IIe (I) or basic Lys (K) at residue 100. Try (W) a bulky hydrophobic amino acid is replaced with the small hydrophilic Gly (G) or basic Arg (R) at residue 122 (Table 3.4.2 & figure 3.22B). These substitutions could all affect the secondary structure of the pre-S domain.

In this study, deletion mutations disrupted the T/B cell pre-S epitopes in at least 48% of the isolates from these HCCs. These selective mutations probably allow for immune evasion. In addition to the structural and functional significance of the pre-S1 and pre-S2 regions, these also appear to be of immunological importance as B and T cell epitopes have been mapped to this region (Milich et al., 1986; Milich et al., 1987; Chisari and Ferrari, 1995). Anti-pre-S2 antibodies were detected in HBsAg/anti-HBs antigen positive patients recovering from HBV infection, but were not detected in chronic carriers (Neurath et al., 1985; Budkowska et al., 1986; Budkowska et al., 1992), with liver cirrhosis or liver cancer (Coursaget et al., 1988). . Experimentally challenged chimpanzees also gain immunity when provided with anti-pre-S1 (Neurath et al., 1989) and anti-pre-S2 antibodies (Neurath et al., 1986a).

4.4.3. PRE-S1/S2 MUTANTS AND GGHS

The role of pre-S gene mutants is further cemented in the development HCC by the characterization of integrated mutant pre-S genes within clonally propagated GGHs in chronic HBV carriers and in HCC patients (Fan et al., 2001; Wang et al., 2003; Hsieh et al., 2004a). These cells are a histological hallmark of HCC and have a glassy appearance because of the abnormal accumulation of envelope proteins within the ER (Wang et al., 2003). Type I GGHs usually contain a mutant pre-S1 gene with a deletion, which interrupts the S2 promoter, and would express mutant LHBs (∆pre-S1LHBs) (section 3.4.2 & figure 3.23). Type I GGHs are usually scattered throughout the liver lobules. Type II GGHs are characterised by pre-S2 internal deletions accompanied by a pre-S2 initiation codon mutation, and would express mutant pre-S2 LHBs (∆pre-S2LHBs) (section 3.4.2 & figure 3.23). Type II GGHs cluster into groups, these mutants usually emerge during late low HBV DNA phase of chronic HBV infection and persist (Fan et al., 2001).

4.4.4. NOVEL TRANSACTIVATORS

$LHBs^{t}$ and $MHBs^{t}$ mutants were identified in 5 (27%) of the HCC patients (section 3.4.3 & figure 3.24). Patient 18 contained $LHBs^{t70}$ and $LHBs^{t124}$/ $MHBs^{t124}$; patient 9 contained $LHBs^{t254}$; patient 26 contained $LHBs^{t271}$, patient 27 contained $LHBs^{t218}$/ $MHBs^{t218}$ and patient 3 contained $LHBs^{t227}$/ $MHBs^{t227}$.

LHBs and $MHBs^{t}$ belong to a family of pre-S2 transactivator proteins. It has been established that the pre-S2-dependent activation of PKC activates downstream targets AP1 and NF-κB via c-Raf-1/Erk2 signal transduction pathway which has been shown to increase hepatocyte proliferation (Hildt et al., 1996a; Hildt et al., 2002). In addition, LHBs and $MHBs^{t76}$ transgenic mouse models have been shown

to develop HCC (Chisari et al., 1989). Furthermore, these C-terminally truncated pre-S/S genes have been described in integrated HBV in several HCC studies (Nagaya et al., 1987; Caselmann et al., 1990; Kekule et al., 1990), with one study identifying these mutants in 25% of the HCC cases (Schluter et al., 1994).

Several factors are required for transcriptional activator function of pre-S2. Their cytosolic topology and retention in the ER secretory pathway or their expression in cytosol is essential (Hildt et al., 1993). In the case of MHBs, at least amino acids 224-281 within the C-terminal hydrophobic region have to be removed by deletion or stop codon mutation (figure 3.24) (Lauer et al., 1992). The pre-S2-dependent activation of PKC requires the structural integrity of the serine 28 phosphorylation site within the cluster PAGG*SSS*G*T*LNP, residues 142-153 of LHB or 23-34 of MHB (figure3.22B) (Hildt et al., 2002).

The carboxyl-truncation positions can also be grouped in relation to the transmembrane hydrophobic domains illustrated in figure 1.5 and figure 3.24. The N-terminal truncation at residue 70 occurs within the first hydrophobic transmembrane domain (figure 3.24). The truncation at residue 124 occurs between the first and second hydrophobic domains. The stop codon 218 causes the exclusion of the third hydrophobic domain, whereas the remaining three 227, 254 and 271 occurs within the third hydrophobic domain. According to the requirements for transcriptional activation of pre-S2, at least the last 70 amino acids of the third hydrophobic domain have to be excluded by deletion or nonsense mutations. According to this criterion, stop codon mutations at residues 70, 124 and 218 should yield active pre-S2 domains while those at positions 227, 254 and 271 would not. Presumably to be active these transactivators have a

cytosolic topology in the ER membrane and are unglycosylated. This has been demonstrated for an MHB mutant truncated at residue 76 ($MHBs^{t76}$) it is ER membrane-bound, cytosolic topology and unglycosylated (Hildt et al., 1993). While $MHBs^{t52}$ is the minimal requirement for transcriptional activation (Hildt et al., 2002), both $MHBs^{t52}$ and $MHBs^{t63}$ are non-membrane bound and have a homogenous distribution in the cell/cytosol (Hildt et al., 1995).

Pre-S2 activation of PKC is also dependent on the structural integrity of S28 (highlighted in figure 3.22B) within the S/T cluster at residues 27, 28, 29 and 31. This cluster is very well conserved in all the isolates (figure 3.22B).

It has been established that LHBs and $MHBs^{t}$ are transcriptional activators (Kekule et al., 1990; Hildt et al., 1993; Hildt et al., 1996a; Hildt et al., 2002). These data have revealed four types of truncation mutants. Type i and ii (section 3.4.3 & figure 3.24) would not express $MHBs^{t}$ due to the deletion of the pre-S2 initiation codon. $LHBs^{t}$ would be expressed from all four types, but their capability as transcriptional activators has not yet been clarified. The transcriptional activity of similar $MHBs^{t}$ expressed from type iii and iv in this study has been investigated in the literature (Lauer et al., 1992; Hildt et al., 1993). A series of $MHBs^{t}$ variants expressed from pre-S2/S genes, truncated between hydrophobic domains I and II (figure 3.24) were shown to transactivate the simian virus 2 promoter when co-expressed with pSV2CAT (Lauer et al., 1992).

4.4.5. CELL CYCLE REGULATORY PATHWAYS AFFECTED BY ∆PRE-SLHBS/MHBS

LHBs and MHBs^{t} have been shown to contribute to HCC by directly interacting with PKC and inducing the ERK pathway (Hildt et al., 1996a; Hildt et al., 2002). Several other pathways are affected by pre-S mutant proteins in an ER-stress independent and/or dependent manner and have been reviewed (Wang et al., 2006).

The ER chaperones, GRP78 and GRP94, are induced when hepatocytes experience ER stress caused by the accumulation of pre-S mutant proteins (Xu and Yen, 1996; Wang et al., 2003). GRP78 gene expression is also upregulated in HCC and GGHs (Types I & II) (Wang et al., 2003; Wang et al., 2005). Furthermore, GRP78 may be involved in anti-apoptotic pathways by interfering with the activity of caspases (Reddy et al., 2003).

Enhanced *Cox*-2 gene expression has been associated with HBV-induced HCC (Cheng et al., 2004). This is supported by studies in hepatoma cell culture models and transgenic mice expressing pre-S mutants (Hung et al., 2004b). Cox-2 specifically enhances tumour angiogenesis (Cheng et al., 2004) and hepatocyte transformation (Hung et al., 2004b).

LHBs containing pre-S2 internal deletions and/or initiation codon mutations (∆pre-S2LHBs) have been shown to directly interact with cellular factors, which affect cell cycle regulation. ∆pre-S2LHBs negatively regulates the activity of tumour suppressor RB (Hsieh et al., 2007) and enhances the expression of cyclin A. Enhanced *cyclin A* gene expression has been demonstrated in HuH7 cells,

transgenic mice (Wang et al., 2005) and in HCC (Ohashi et al., 2001). Dysregulation of cyclin A expression could not only affect cell proliferation directly, but may also affect DNA replication and centrosome duplication as cyclin A is involved in these functions (Meraldi et al., 1999). Pre-S mutants induce ER stress related oxidative DNA damage in HuH 7 cells (Hsieh et al., 2004a). Furthermore, the base excision repair (BER) pathway is induced in the ∆pre-S2LHBs transgenic mouse model and in human HBV-induced HCCs (Hsieh et al., 2004a). These data indicate that genome integrity can be compromised in liver tissues expressing mutant LHBs, which could contribute to hepatocarcinogenesis.

4.4.6. CONCLUSION

The detailed analysis of 47 pre-S/S gene isolates from HCC patients and their comparison with 23 pre-S gene isolates also of southern Africa origin from individuals with different clinical manifestations has revealed several findings.

The pre-S deletion and initiation codon mutations were prevalent in strains isolated from HCC patients compared with those present in ASC, AC and FH. This finding is supported by several studies which show a very strong correlation between pre-S mutants and the development of HCC (Takahashi et al., 1998; Huy et al., 2003; Sugauchi et al., 2003a). This is the first study, which describes the pre-S mutants in HCC in southern Africans, it confirms that these mutants are also prevalent in subgenotype A1 HBV and probably contribute to HCC in the patients, which harbour this subgenotype.

The molecular characterization of the pre-S mutants revealed that initiation codon and deletion mutations are more prevalent in the pre-S2 gene compared to the pre-S1 gene, in the HBV strains isolated from HCCs. It has also been established

previously that both mutants lead to the development of GGHs with different histopathological characteristics (Fan et al., 2000; Wang et al., 2003). Data are also accumulating to suggest that ∆pre-S2LHBs and ∆pre-S1LHBs may affect their pathogenesis differently, by their differential interaction with cellular proteins. In addition to identifying the mutant genes encoding the well characterised ∆pre-S1LHBs, ∆pre-S2LHBs, ∆pre-S1pre-S2LHBs and MHBs^t the novel LHBs^t and ∆pre-S1pre-S2LHBs^t were characterised. Recent data have implicated these proteins as direct transactivators of key cell regulatory proteins or in the induction of several pathways in an ER-stress related manner, reviewed by (Wang et al., 2006).

4.5. MUTANT PRE-S/S GENES AND THE CELL CYCLE REGULATORY PATHWAYS

The expression of pre-S mutants leads to the intracellular retention of HBV envelope proteins, which is a major cause of ER stress leading to the activation of several downstream pathways, which eventually affect cell survival, proliferation and genome integrity (reviewed in (Wang et al., 2006). ER stress induced by pre-S mutants generates reactive oxygen species (ROS), which causes DNA damage (Hsieh et al., 2004a). DNA damage can in turn lead to the activation of cell cycle regulatory pathways: cell cycle arrest (senescence) and apoptosis (Hsieh et al., 2004c; Hampton, 2000). Various factors such as viral proteins can also simultaneously disrupt these pathways (Ahn et al., 2001).

The p53 tumour suppressor protein is a key regulator of these cellular pathways (Levine, 1997; Vogelstein et al., 2000a) and responds to numerous cellular

stresses by executing the transcriptional activation of down stream regulatory genes such as $p21^{Waf-1}$ (Vogelstein et al., 2000b).

Early studies suggested that the $p21^{Waf-1}$ - p53 pathway may be disrupted in HBV-associated HCC. $P21^{Waf-1}$ expression levels are lower in HBV chronic carriers compared to HCV (Wagayama et al., 2001). Lee et al also showed a discordant relationship between p53 and $p21^{Waf-1}$ expression in HBV-associated HCC by immunohistochemical analysis (Lee et al., 2004). They reasoned that increased p53 expression as a result of aberrant loss of function, disrupted p53 dependent $p21^{Waf-1}$ activation and resulted in decreased $p21^{Waf-1}$ protein in HCC. Hepatoma cell lines, which contain integrated HBV, PLC/PRF/5 and Hep G/2215 also present with p53 anomalies (Puisieux et al., 1993).

As a pilot screen to determine whether mutated pre-S/S genes and their relevant mutated envelope proteins, especially Δpre-S1/pre-S2LHBs^{t}, could affect these cell regulatory pathways and whether they warrant further investigation their effect was determined on $p21^{Waf1}$ protein levels and caspase 3 activity in HepG2 cells. HepG2 cells were chosen as the preferred *in vitro* model because these cells closely match the biochemical and phenotypic characteristics of differentiated hepatocytes (Knowles et al., 1984). Doxorubicin treatment successfully activated both the senescence ($p21^{Waf1}$) and the apoptotic (caspase 3) pathways within the HepG2 cell line (figure 3.27 C & D).

Overall it appears that the transfection with all five pre-S mutant gene constructs was able to induce a decrease in $p21^{Waf1}$ protein expression, and this may be independent of caspase 3 (figure 3.27 C &D). During apoptosis, $p21^{Waf1}$ has been

shown to undergo proteolytic cleavage by caspase 3 in colorectal cancer cells (Jin et al., 2000) and in SK Hep-1 human hepatoma cells (Tang et al., 2006) leading to its depletion. However, it may be more informative to measure apoptosis at later time points or in a stable mammalian cell line containing inducible mutant pre-S/S genes.

Pre-S mutant 9.5, which encodes LHBs with a large pre-S1/pre-S2 internal deletion and 3' truncation (Δpre-S1/pre-S2LHBs^{t}) (figure 3.22 A, B & 3.26) appears to induce the most severe reduction in p21^{Waf1} levels, followed by 12.9, 2.2 and 1.2 (figure 3.27C). The latter constructs contain a combination of pre-S2 initiation codon and/or internal deletions (Δpre-S2LHBs) known to drastically reduce MHBs and SHBs expression and induce ER stress. In pre-S mutant 9.5, it is not clear whether the combination of a pre-S1/pre-S2 internal deletion and a 3' truncation would confer additional transactivational capacity or whether its effect on p21 ^{Waf1} is mediated through ER stress. Evidence is steadily accumulating for the ER stress-independent induction of cell cycle regulatory factors by pre-S mutants. Δpre-S2LHBs have been shown to upregulate cyclin A expression in HuH7 cells and liver tissues (Ohashi et al., 2001; Wang et al., 2005). These mutants could also improve the deviant transformation characteristics of non transformed hepatocytes and induce nodular hepatocyte proliferation in transgenic mice (Wang et al., 2005). The Δpre-S2LHBs are also able to induce the hyperphosphorylation of tumour suppressor RB by directly interacting with JAB1 (Hsieh et al., 2007).

While a direct disruption of the $p21^{Waf-1}$ - p53 pathway has not been described by mutant envelope proteins, several *in vitro* analyses of wild-type HBx has established that this protein is able to disrupt the $p21^{Waf-1}$ - p53 pathway. HBx has been shown to directly interact with the p53 protein via its 102-134 C-terminal domain and execute cytosolic sequestration and prevent p53-mediated apoptosis in primary hepatocytes (Elmore et al., 1997). HBx was also shown to interfere with *$p21^{Waf-1}$* gene promoter activity by downregulating Sp1 activity and decreasing $p21^{Waf-1}$ transcription (Ahn et al., 2001). HBx may indeed have a role to play in HCC by disrupting p53 function and decreasing $p21^{Waf-1}$ expression. However, C-terminal mutations are commonly reported in the X gene in HBV in HCC and these mutations abrogate transactivational domains (Poussin et al., 1999; Sirma et al., 1999; Wang et al., 2004). On the other hand, pre-S/S C-terminal mutations serve to activate pre-S transactivators and have been reported prolifically in HCC (Kekule et al., 1990; Santantonio et al., 1992; Huy et al., 2003; Sugauchi et al., 2003a). The role of the mutant pre-S/S proteins as the preferred HBV transactivators is also supported by the direct comparison of integrated pre-S/S and X gene (Caselmann et al., 1990).

In this study one patient (1/9 in complete genome data set) contained five strains with X gene premature truncation mutations and two strains also had the X gene initiation codon abolished. These mutations excluded the residues 102-136 required for p53 interaction. The X-gene was deleted in 2/4 complete genomes identified in Korean HCC patient (Kim et al., 2007). This is supported by several studies, which show that common C-terminal X-gene associated with HCC may abrogate the anti-proliferative properties of HBx (Sirma et al., 1999; Tu et al.,

2001). Whereas, the majority of HCC patients contain HBV strains which possess pre-S/S mutations (Ohno et al., 1993; Sugauchi et al., 2003a). These pre-S/S mutants have been implicated in cellular transformation and proliferation (Cheng et al., 2004; Hsieh et al., 2004a; Hung et al., 2004b; Hsieh et al., 2007).

In the current study 72% of the HCC patients contain HBV strains which possess pre-S/S mutations. It is feasible to suggest that these mutants could also contribute to hepatocarcinogenesis by disrupting the $p21^{Waf-1}$ levels in a p53 dependent or independent manner. ∆pre-S2LHBs and ∆pre-S1/pre-S2LHBs^{t} may prove to directly inhibit $p21^{Waf-1}$ similarly to HBx by affecting transcriptional regulation of this gene.

Interestingly enhanced p53-mediated apoptosis was observed in COX-2 null mammalian cells, this suggested that COX-2 may have a negative effect on p53 mediate apoptosis (Han et al., 2002). Enhanced expression of COX-2 as observed in ∆pre-S1LHBs and ∆pre-S2LHBs expressing cells and tissues (Hung et al., 2004b) may lead to the dysregulation of p53 activities.

ER-stress induced p53 degradation is not entirely ruled out since it was shown that when either physiological or pharmacological ER stress was induced in mammalian cells (Hung et al., 2004a), this could lead to the inhibition of p53 (Qu et al., 2004). ER stress was able to induce glycogen synthase kinase-3 β (GSK-β) binding to p53 and this led to its cytoplasmic translocation and subsequent degradation.

4.5.1. CONCLUSION

Evidence is accumulating which indicate that mutant pre-S/S genes and their relevant mutant MHBs/LHBs can directly or indirectly abrogate the function of key cell cycle regulators such as cyclin A, RB and and COX-2 (Hung et al., 2004b; Wang et al., 2005; Hsieh et al., 2007). In this study, several novel mutant pre-S/S genes have been described and their function should be evaluated further. The variants may reduce $p21^{Waf1}$ levels and disrupt cell cycle control and in this way contribute to hepatocyte transformation. However, further experiments would have to be done to support these observations. The initial analysis would have to involve determining the effects on p53 and $p21^{Waf1}$ transcript levels to determine which possible mechanisms may be involved.

CHAPTER 5

5.0 GENERAL CONCLUSION

It has been over three decades since the discovery of HBV in the 1960's, when B.S. Blumberg discovered the Australian antigen. Since then HBV has been studied vigorously and enormous progress has been made in understanding its molecular biology, evolution and disease aetiology. Despite the development of a successful vaccine, more 360 million people remain chronically infected with this virus and are at risk of developing HCC, and a substantial number of these individuals reside in Africa. Despite the high HBV endemnicity in Africa there is a paucity in the data from this continent. Sequence data on complete HBV genomes isolated from African HCC patients are particularly rare. The current study has successfully characterised complete HBV genomes from southern African black HCC patients and has made several novel findings.

The role of the 1762T1764A mutations and 1809 - 1812 variants are well characterised in HCC and subgenotype A1 in southern Africa, respectively (Baptista et al., 1999; Ahn et al., 2003; Kimbi et al., 2004). In the current study an expanded analysis of the BCP and pre-C has revealed high replicating variants containing combinations of the 1753C1762T1764A1766T1768A1809T1812T mutations within single genomes, which would also express reduced HBeAg levels. Previous studies have shown that these mutations regulate viral replication and gene expression at various levels: transcriptional (1762T1764A) (Buckwold et al., 1996), post-transcriptional and/or translational (1766T1768A) (Baumert et al., 1998) and translational (1809 - 1812) (Ahn et al., 2003).

Furthermore, the BCP and pre-C variants were found in combination with pre-S/S gene mutations within single genomes. By implication these enhanced replicative variants would result in the accumulation of replicative intermediates, and the expression of dysfuncational viral proteins and transactivators. These characteristics could initiate and contribute to disease by encouraging viral integration (Raimondo et al., 1988), cellular stress due to intracellular accumulation of proteins and dysregulation of cell cycle pathways.

Two patients did not contain HBV strains with the double BCP mutation (patient 6 and patient 18). However, all the isolates identified in patient 18 possessed a novel complex BCP mutation which has not been described in HBV strains isolated from South Africa previously. The BCP contained a 13 nt deletion (nt 1752 – 1764) and a 45 nt insertion at position 1770 (figures 3.9, 3.10 and 3.11). These mutations abrogated the pre-C promoter and introduced two HNF1 transcription factor binding sites and 1 putative HNF3, creating high virus replication variant with reduced HBeAg expression. Patient 6 contained HBV strains with the single BCP mutation 1764A**.** This means that none of the complete genomes described in these HCC patients were spared from BCP alterations and this evidence supports the correlation between BCP mutations and the development of HCC.

Phylogenetic analysis exposed three ASC that may be at risk for developing HCC. Not only did the isolates from these ASC consistently group with HCC sequences within the dendograms of all the ORFs but also shared the BCP 1762T1764A mutations. This is supported by previous analyses, which show a significant

association between the double BCP mutation and HCC in South Africa (Baptista et al., 1999).

To our knowledge this is the first study to isolate and characterise a splice variant and a poly (dA) variant from the serum of a black southern African patient infected with subgenotype A1. Splice variants do not appear to be essential for HBV replication, but may indirectly be involved in disease through the enhanced expression and intracellular accumulation of viral proteins. Furthermore, there appears to be evidence for the intracellular accumulation and integration of spliced genomes. This knowledge could be useful in the pursuit of characterising HBV integrants.

The role of the poly (dA) variant in HBV-related disease is not clear. It has been postulated that this defective genome is produced in hepatocytes by an unconventional replication strategy and encapsidated and secreted into the serum (Sommer et al., 1997). The possible expression of HBx and defective viral proteins could contribute to pathogenesis.

This study also presents a comprehensive analysis of the pre-S mutants in HCC in southern Africans, it confirms that these mutants are also prevalent in subgenotype A1 HBV and probably contribute to HCC in the patients which harbour this subgenotype. Several studies show a very strong correlation between pre-S mutants and the development of HCC (Takahashi et al., 1998; Huy et al., 2003; Sugauchi et al., 2003a). The molecular characterization of the mutant pre-S/S revealed that initiation codon and deletion mutations are more prevalent in the pre-S2 gene compared to the pre-S1 gene, in the HBV strains isolated from HCCs. This is probably because the pre-S2 domain contains B - and T- cell immune

epitopes and experiences severe selection pressure during chronic infection. Data are also accumulating to suggest that ∆pre-S2LHBs and ∆pre-S1LHBs may affect their pathogenesis differently, by their differential interaction with cellular proteins.

In addition to identifying the mutant genes encoding the well characterised ∆pre-S2LHBs, ∆pre-S1LHBs, ∆pre-S1pre-S2LHBs and MHBs^t, the novel LHBs^t and ∆pre-S1pre-S2LHBs^t were characterised. Recent data have implicated these proteins as direct transactivators of key cell regulatory proteins or in the induction of several pathways in an ER-stress related manner.

Despite the anticipated intra- and inter-patient HBV heterogeneity there appears to be several common characteristics shared between the emerging variants. These are primarily the accumulation of BCP and/or pre-C mutations (1753C1762T1764A1766T1768A1809T1812T) as well as pre-S/S mutations. These probably lead to the accumulation of replicative intermediates and viral proteins, contributing to viral integration and cellular stress or damage. In addition, the expression of viral transactivators could contribute to cellular transformation. These combined characteristics could progressively induce liver disease and HCC and should be explored further where necessary and be considered during therapeutic design.

APPENDICES

APPENDIX A:

A. Buffers and solutions

A1.1. Tris-Borate EDTA (TBE) pH 8.3
0.89 M Tris base
0.89 M Boric Acid
25 mM EDTA

A1.2. Tris-Acetate EDTA (TAE)
0.04 M Tris-acetate
0.002 M EDTA (pH 8.0)

A2.1. SOC Medium pH 7.0
2 g Tryptone
0.5 g Yeast Extract
1 ml 1 M NaCl
0.25 ml 1 M KCl
1 ml 2 M Mg^{2+} stock (1 M $MgCl_2 \cdot 6H_2O$, 1 M $MgSO_4 \cdot 7H_2O$)
1 ml 2 M glucose
Filter sterilize

A2.2. Luria Broth (LB) pH 7.5

Tryptone	10	g
Yeast Extract	5	g
NaCl	10	g

A2.3. Transformation Buffer A pH 7.0

Glycerol	15	ml
$CaCl_2$.2H2O	1.47	g
Pipes	0.3	g
Final volume	100	ml

A2.4. X-gal
5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal)
50mg/ml in dimethylformamide stored at -20℃.

A2.5. Ampicillin
100 mg/ml in deionised water (filter sterilize)

A2.6. IPTG stock solution (0.5 M)
6 g isopropyl β-D-thiogalactopyranoside (IPTG) in 50 ml
of deionised water. Filter sterilized with a 0.2 µm filter store at -20℃.

A3. Manual Sequencing
A3.1. 40% Acrylamide (Sequencing)
380 g Acrylamide
20 g Bisacrylamide

Dissolve in 1 L of ddH2O
Deionised with 50 g of ion exchanger overnight at 4℃

A3.2. Glycerol Tolerant Buffer (GTB) 50 ml
108 g Tris-Base
36 g Taurine
2 g EDTA

A3.3. 8 % Sequencing Gel 120 ml
50.4 g Urea
6 ml GTB
24 ml 40 % Acrylamide

A4. 10 X Phosphate Buffered Saline (PBS) ph 7.4
1.37 M NaCl
26.8 mM KCl
43 mM Na_2HPO_4
14.7 mM KH_2PO_4

A5. 50 mM HEPES pH 7.1
280 mM NaCl
1.5 mM Na_2HPO_4

A6. 2 X HEPES Buffer Saline (HBS) pH 7.1
50 mM HEPES, pH 7.1
280 mM NaCl
1.5 mM Na_2HPO_4

A7. Mycoplasma Test

A7.1Stain
0.5 µg/ml Hoeschst No. 33258 in Hanks balanced salt solution
(without phenol red or Sodium Bicarbonate)
Stored covered bottle at 4℃, check regularly for contamination.

A7.2 Fixative
1:3 glacial acetic acid : methanol

A7.3 Mounting Fluid
0.0222 M Citric Acid
0.0556 M Na2HPO4.2H2O
50 % Glycerol
pH 5.5 critical for optimal fluorescence
store at 4℃

A7.4 Fixative
4 % paraformaldehyde dissolve 0.2 g in 5 ml of PBS at 60℃.
Allow to cool at room temperature before using.
Aliquot and store at -20℃. Use only once after thawing.

A8 β-Galactosidase
X-gal
5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal)
50 mg/ml in dimethylformamide stored at -20℃.

A8.1 X-gal stain
20 mM Potassium Ferrocyanide mw 329.26
20 mM Potassium Ferricyanide mw 422.4
2 mM MgCl2
Makeup in 1 X PBS pH7.4

200 µl 100 mM K_3 Fe $(CN)_6$
200 µl 100 mM K_4 Fe $(CN)_6$
2 µl 1 M $MgCl_2$
598 µl 1 X PBS ph 7.4 (endogenous β-gal active at pH 6.0)
20 µl 50 mg/ml X-gal

A9. Western Blot

A9.1. RIPA Buffer
150 mM NaCl
1% Triton X 100
10 mM Tris pH 7.5
1% Deoxycholate

A9.2. 30% Acrylamide
30 g Acrylamide
0.8 g Bisacrylamide
0.1 g SDS

A9.3. Resolving Gel Buffer 200 ml pH 8.9 4℃

36.2 g Tris

0.8 g SDS

A9.4. Stacking Gel Buffer 100 ml pH 8.0 4℃
5.9 g Tris
0.4 g SDS

A9.5. 5X Loading Buffer 50 ml pH 6.8
1.75 g Tris
30 ml Glycerol
5g SDS

A9.6. 5X Loading dye
200 µl 5X loading buffer
100 µl β-mercaptoethanol
100 µl Bromophenol Blue (saturated solution)

A9.7. 10X Running Buffer 1L
63.2 g Tris
40 g Glycine
10 g SDS

A9.8. 10 X Transfer Buffer 1L
144 g Glycine
38 g Tris

A9.9. 10 X Tris Buffered Saline (TBS) 1L pH 7.5
60.5 g Tris
87.6 g NaCl

A9.10. Ponceau S Stain
0.001% Ponceau S
0.01% Glacial Acetic Acid

A9.11. Coomassie Blue Staining Solution
50 % Methanol
10 % Glacial Acetic Acid
0.25 % Coomassie Blue

A9.12. Destaining Solution
0.05 % Methanol
0.07 % Glacial Acetic Acid

A9.13. Stripping Buffer
14.3 M β-mercaptoethanol
10 % SDS
1 M Tris HCl

A10. Caspase Assay
A10.1. Doxorubicin
2 mg/ml prepared in sterile water.

A10.2. Lysis Buffer pH 7.5
10 mM Tris-HCl (pH 7.5)
10 mM NaH_2PO_4
130 mM NaCl
1 % Triton X 100
10 mM NaPPi
filter sterilize with 0.45uM filter

A10.3. Reaction Buffer
100mM HEPES pH 7.5
10% Sucrose
10mM DTT
0.001 % NP40
0.1 % CHAPS pH7.5

A10.4. DEVD-AMC (aspartate-glutamate-valine-aspartate-7-amino-4-methyl coumarin)

1 mg/ml in Dimethyl Sulphoxide

APPENDIX B Figure B1. Nucleotide sequence variation of complete genome isolates from HCC.

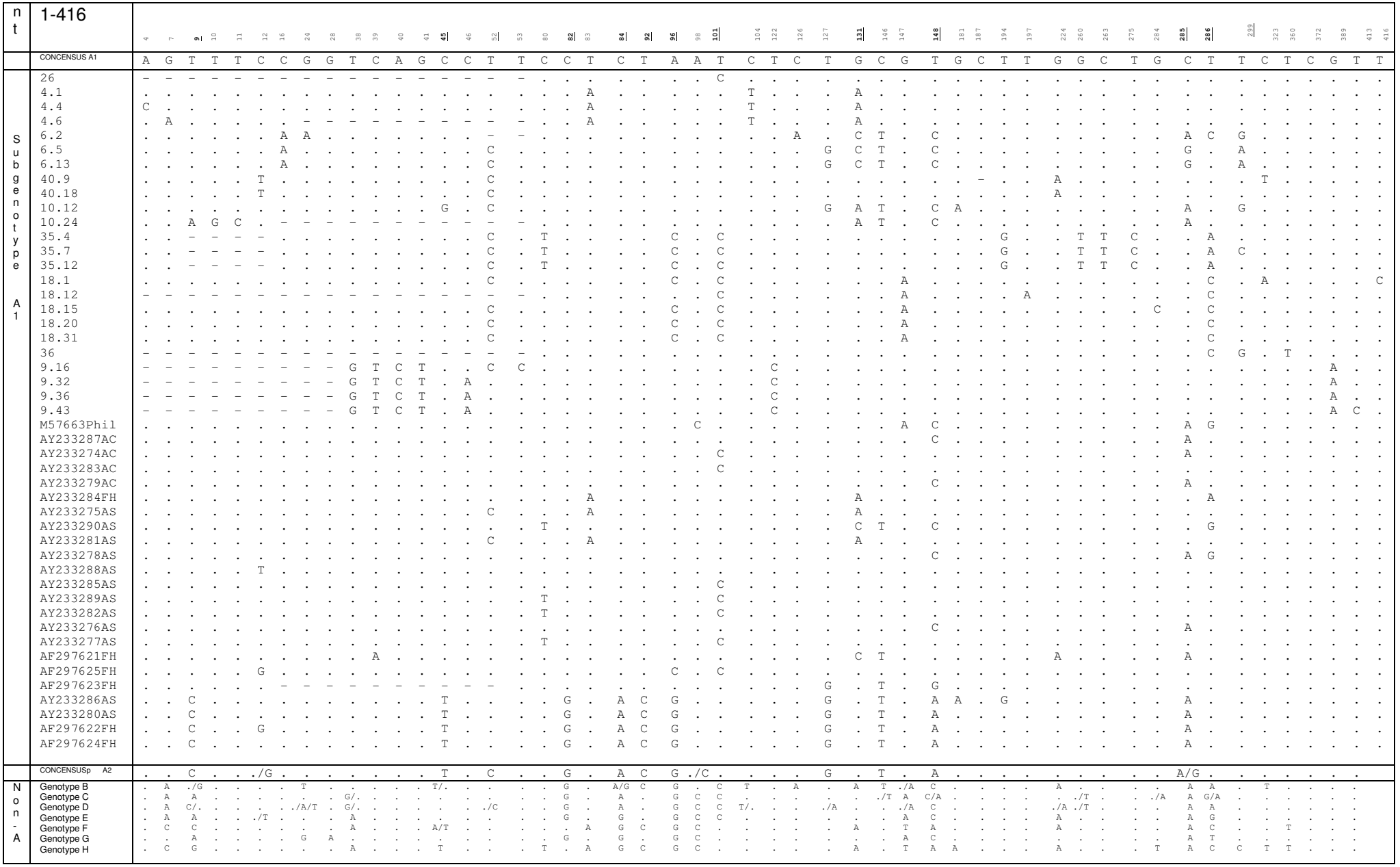


Figure B1 continued

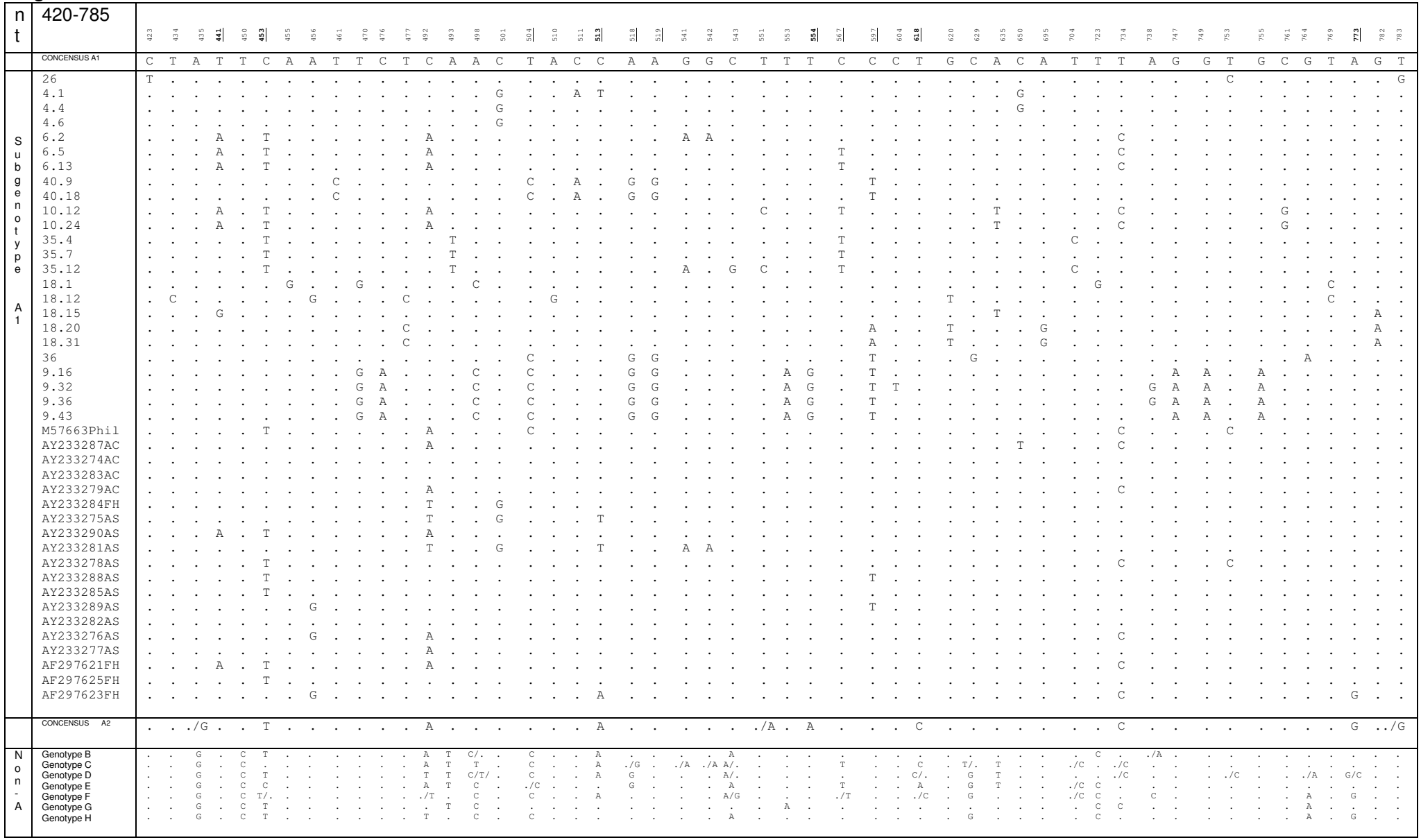

nt
420-785
423 434 435 441 450 453 455 456 461 470 476 477 492 493 498 501 504 510 511 513 518 519 541 542 543 551 553 554 567 597 604 618 620 629 635 650 695 704 723 734 738 747 749 753 755 761 764 769 773 782 783
CONCENSUS A1
C T A T T C A A T T C T C A A C T A C C A A G G C T T T C C C T G C A C A T T T A G G T G C G T A G T
Subgenotype A1
26 T . . .

Figure B1 continued

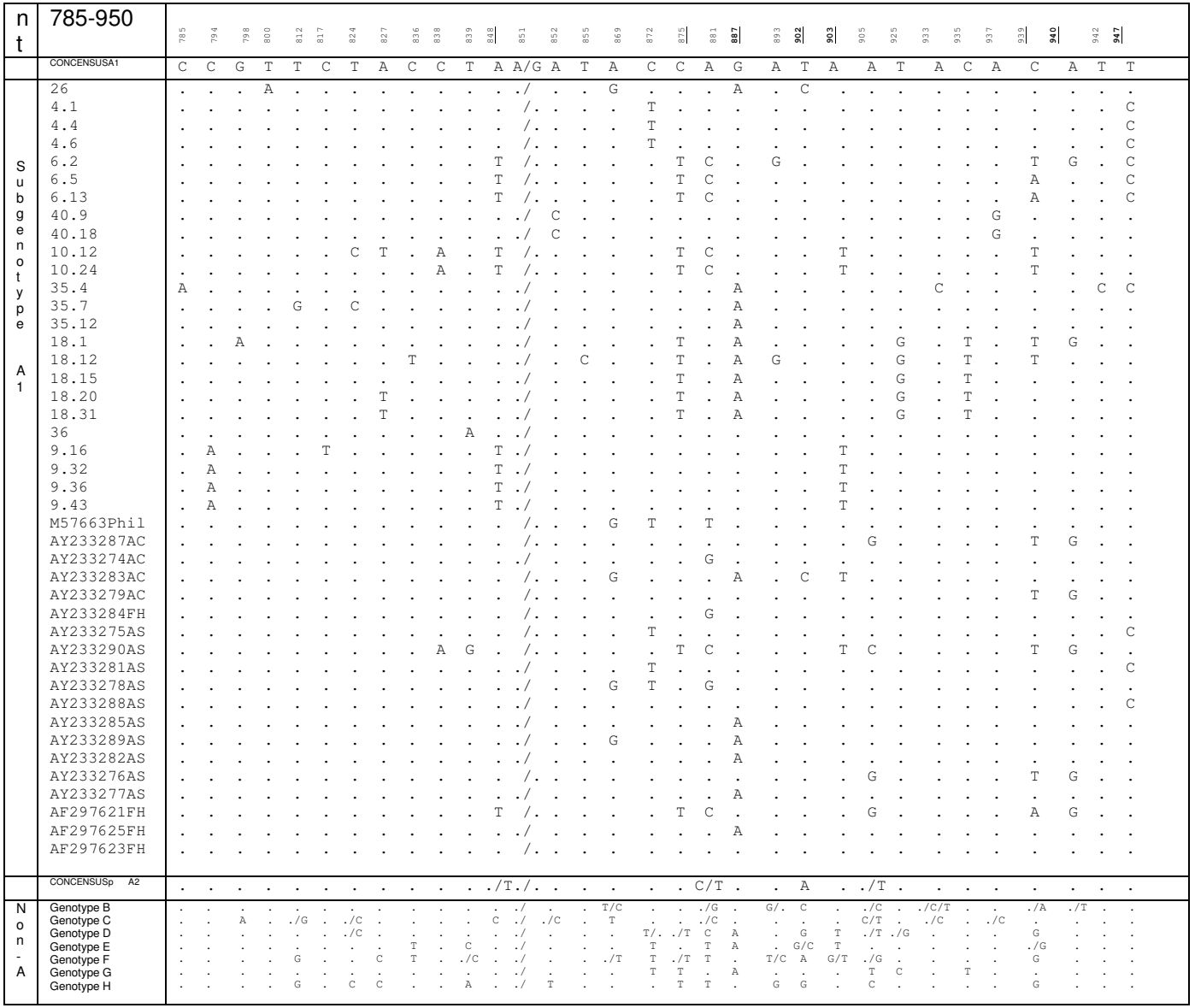


Figure B1 continued

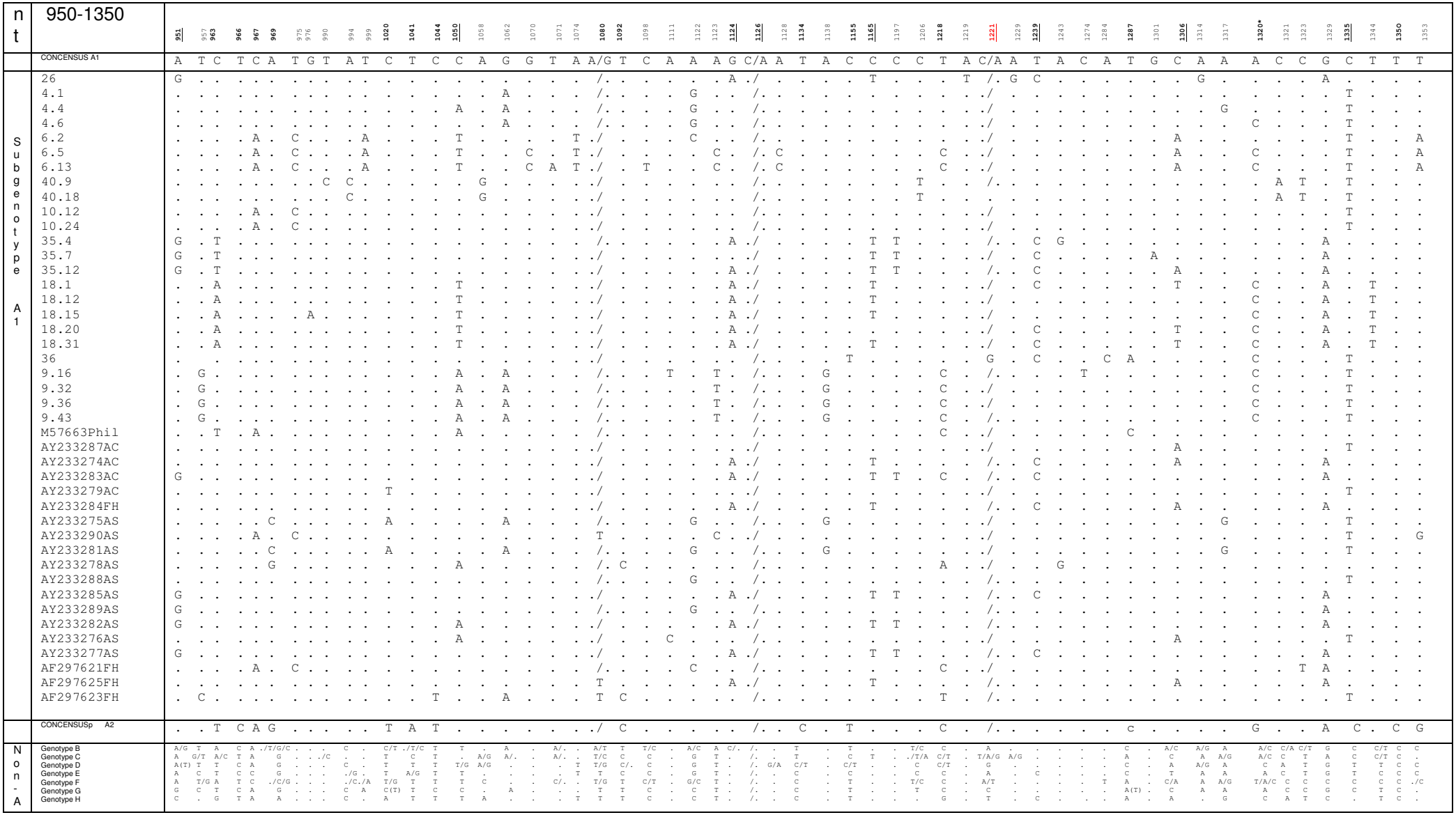


Figure B1 continued

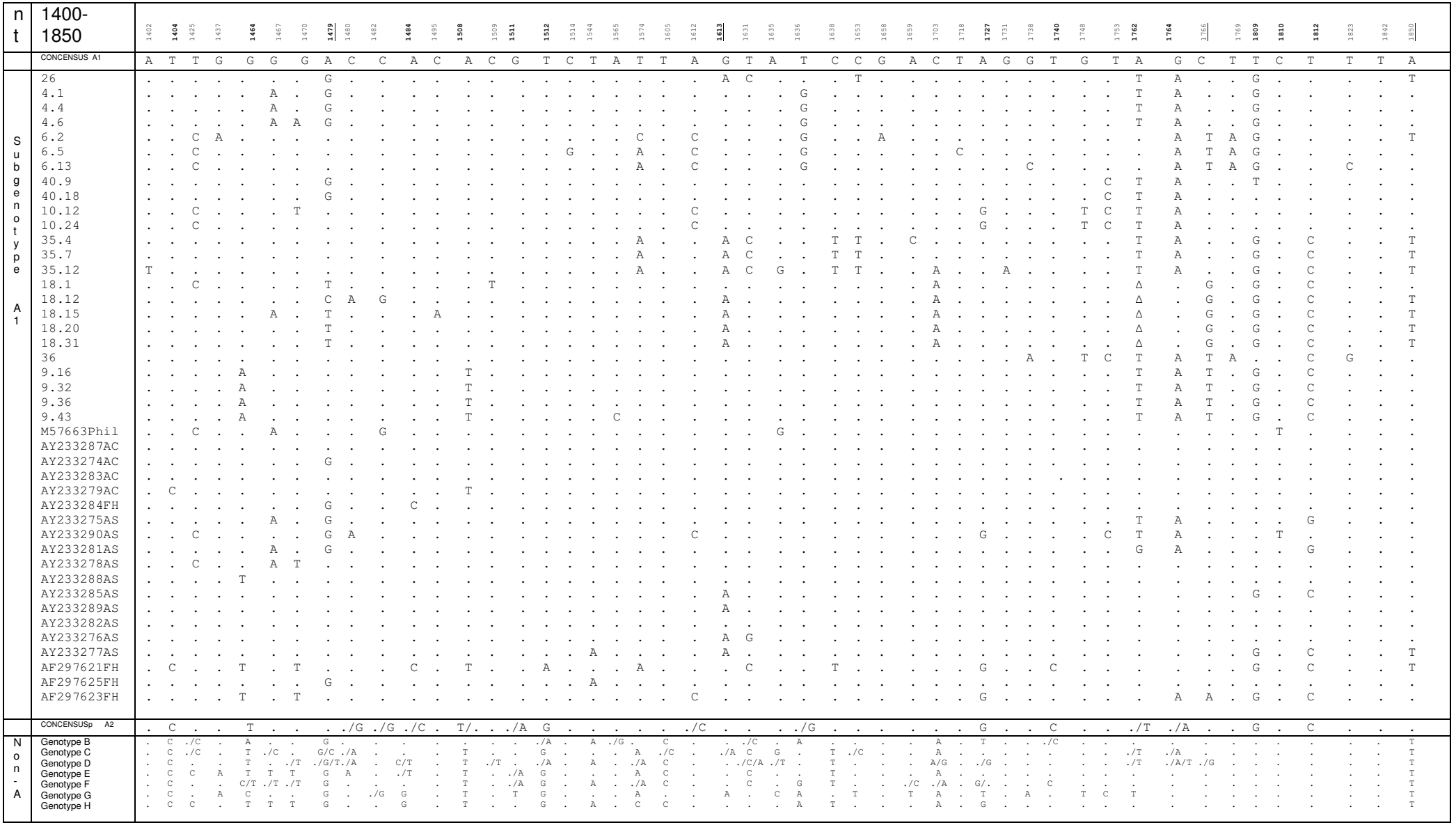


Figure B1 continued

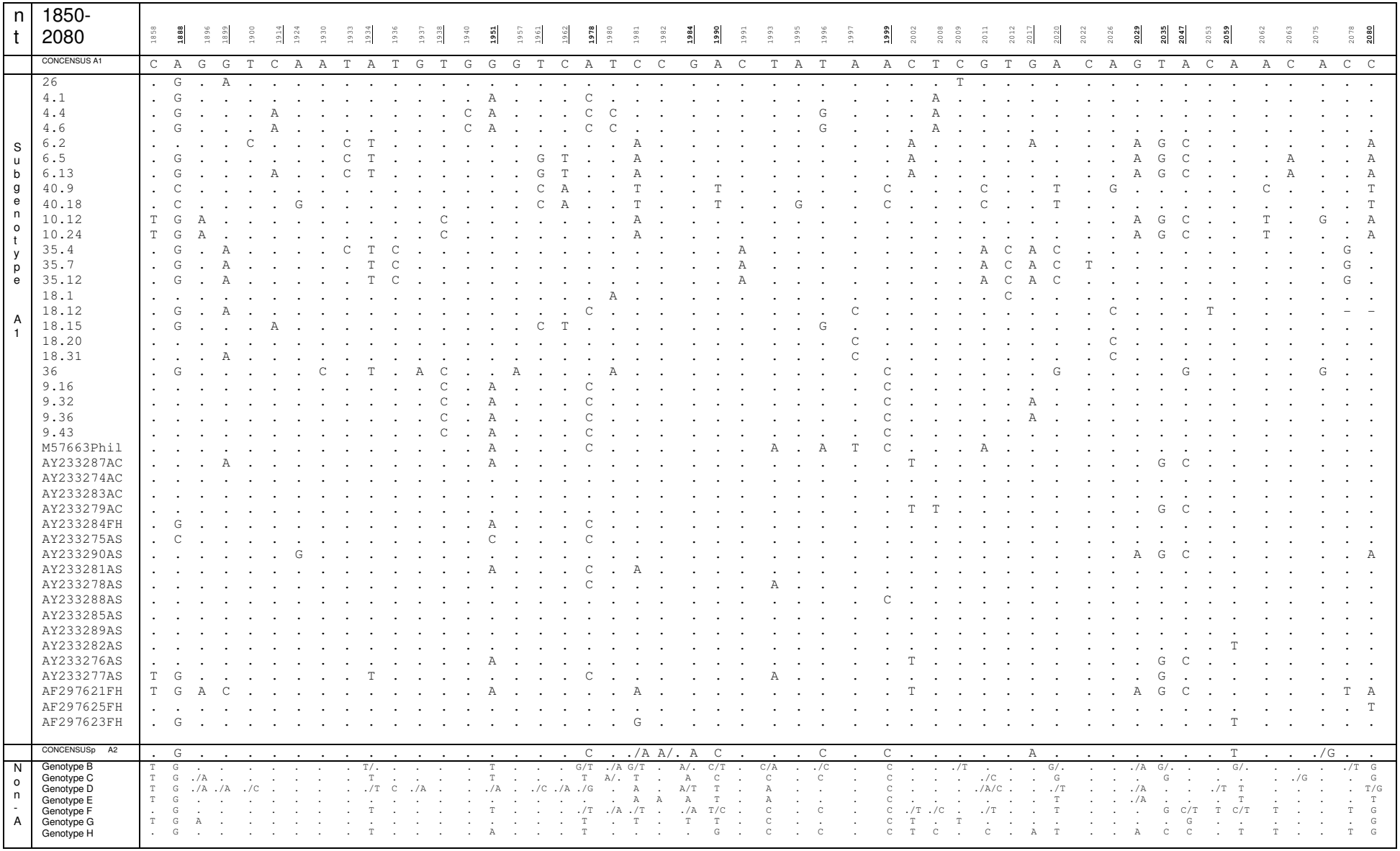


Figure B1 continued

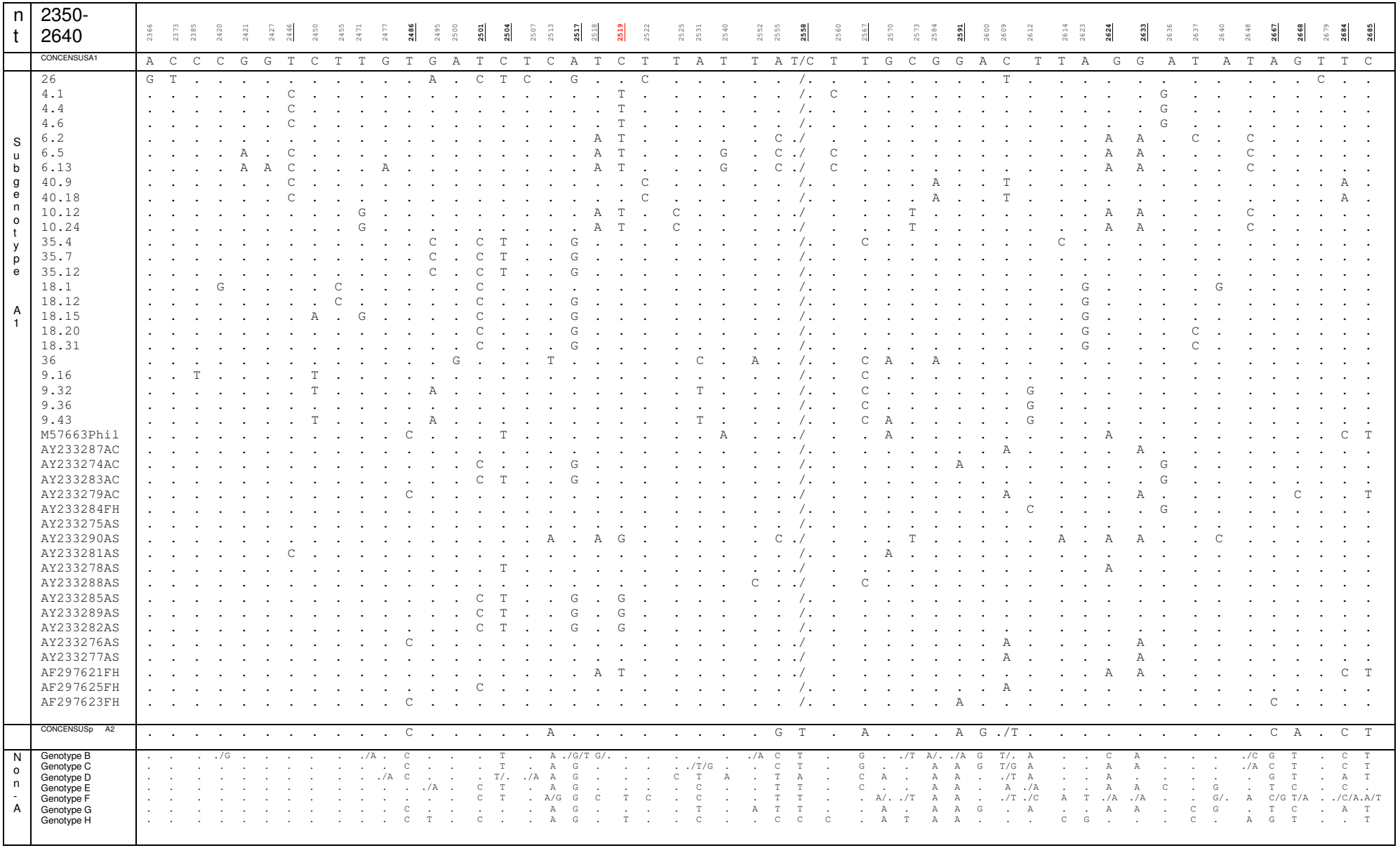


Figure B1 continued

nt	2685-2750	2687	2690	2693	2696	2702	2711	2714	2717	2720	2721	2723	2726	2735	2738	**2741**	2744	**2745**
	CONCENSUSA1	A	C	A	G	A	T	T	T	A	G	A	T	C	C	G	C	C/A
Subgenotype A1	26	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	./
	4.1	.	.	.	.	G	.	.	.	.	.	.	.	.	.	A	.	/.
	4.4	.	.	.	.	G	.	.	.	.	.	.	.	.	.	A	.	/.
	4.6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	/.
	6.2	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	/.
	6.5	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	/.
	6.13	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	/.
	40.9	C	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	./
	40.18	C	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	./
	10.12	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	/.
	10.24	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	/.
	35.7	.	.	.	.	.	.	.	.	.	T	G	.	.	.	.	.	./
	35.4	.	.	.	.	.	.	G	.	G	T	G	.	.	.	.	.	./
	35.12	.	.	.	.	.	.	.	.	.	T	G	.	.	.	.	.	./
	18.1	T	.	.	.	.	.	A	.	.	.	.	.	.	A	.	.	./
	18.12	T	.	.	.	G	.	A	C	G	.	.	.	.	.	.	.	./
	18.15	T	.	.	.	.	C	A	.	.	.	.	.	.	.	.	.	./
	18.20	T	.	.	.	.	.	A	.	.	.	.	.	.	A	.	.	./
	18.31	T	.	.	.	.	.	A	.	.	.	.	.	.	A	.	.	./
	36	.	.	G	.	.	.	.	C	.	.	.	.	.	.	.	.	./
	9.16	.	.	.	.	.	.	.	.	.	.	.	.	T	.	A	A	/.
	9.32	.	.	.	.	.	.	.	.	.	.	.	G	T	.	A	A	./
	9.36	.	.	.	.	.	.	.	.	.	.	.	G	T	.	A	A	./
	9.43	.	.	.	.	.	.	.	.	.	.	.	G	T	.	A	A	./
	M57663Phil	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	./
	AY233287AC	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	/.
	AY233274AC	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	./
	AY233283AC	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	./
	AY233279AC	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	/.
	AY233284FH	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	/.
	AY233275AS	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	/.
	AY233290AS	.	.	.	.	.	.	A	.	.	A	.	.	.	.	.	.	/.
	AY233281AS	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A	.	/.
	AY233278AS	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	./
	AY233288AS	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	./
	AY233285AS	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	./
	AY233289AS	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	./
	AY233282AS	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	./
	AY233276AS	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	/.
	AY233277AS	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	/.
	AF297621FH	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	./
	AF297625FH	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	./
	AF297623FH	.	.	.	A	.	.	.	.	.	A	.	.	.	.	.	.	/.
	CONCENSUSp A2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A
Non-A	Genotype B	.	T	.	./T/A	.	.	C	G	T	.	.	.	.	.	G	G/T	A/C
	Genotype C	G/.	./T	.	C	.	.	./C	A	T/G	.	.	.	.	A/.	A	T/A	A
	Genotype D	G/.	T	G	T	.	.	A/G	A	T	C/T	.	.	.	.	A	./T	A
	Genotype E	.	G	.	T	.	.	A	.	T	.	.	.	.	.	A	.	A
	Genotype F	G	G/T	.	C/.	.	.	A	G/.	T/G	.	.	.	T	./A	A	.	A
	Genotype G	.	.	.	T	.	.	A	A	T	.	.	.	.	.	G	.	A
	Genotype H	G	T	.	T	G	.	A	G	T	.	G	.	.	A	A	T	A

Reference List

1. Ahn,J.Y., Chung,E.Y., Kwun,H.J., and Jang,K.L. (2001). Transcriptional repression of p21(waf1) promoter by hepatitis B virus X protein via a p53-independent pathway
3. Gene *275*, 163-168.

2. Ahn,S.H., Kramvis,A., Kawai,S., Spangenberg,H.C., Li,J., Kimbi,G., Kew,M., Wands,J., and Tong,S. (2003). Sequence variation upstream of precore translation initiation codon reduces hepatitis B virus e antigen production. Gastroenterology *125*, 1370-1378.

3. Alberti,A., Clumeck,N., Collins,S., Gerlich,W., Lundgren,J., Palu,G., Reiss,P., Thiebaut,R., Weiland,O., Yazdanpanah,Y., and Zeuzem,S. (2005). Short statement of the first European Consensus Conference on the treatment of chronic hepatitis B and C in HIV co-infected patients. J. Hepatol. *42*, 615-624.

4. Allain,J.P. (2004). Occult hepatitis B virus infection. Transfus. Clin. Biol. *11*, 18-25.

5. Allen,M.I., Deslauriers,M., Andrews,C.W., Tipples,G.A., Walters,K.A., Tyrrell,D.L., Brown,N., and Condreay,L.D. (1998). Identification and characterization of mutations in hepatitis B virus resistant to lamivudine. Lamivudine Clinical Investigation Group. Hepatology *27*, 1670-1677.

6. Arauz-Ruiz,P., Norder,H., Robertson,B.H., and Magnius,L.O. (2002). Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. J. Gen. Virol. *83*, 2059-2073.

7. arcon-Vargas,D. and Ronai,Z. (2002). p53-Mdm2--the affair that never ends. Carcinogenesis *23*, 541-547.

8. 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.

9. Asare,G.A., Mossanda,K.S., Kew,M.C., Paterson,A.C., Kahler-Venter,C.P., and Siziba,K. (2006). Hepatocellular carcinoma caused by iron overload: a possible mechanism of direct hepatocarcinogenicity
3. Toxicology *219*, 41-52.

10. Bancroft,W.H., Mundon,F.K., and Russell,P.K. (1972). Detection of additional antigenic determinants of hepatitis B antigen. J Immunol. *109*, 842-848.

11. Banerjee,A., Banerjee,S., Chowdhury,A., Santra,A., Chowdhury,S., Roychowdhury,S., Panda,C.K., Bhattacharya,S.K., and Chakravarty,R. (2005). Nucleic acid sequence analysis of basal core

promoter/precore/core region of hepatitis B virus isolated from chronic carriers of the virus from Kolkata, eastern India: low frequency of mutation in the precore region. Intervirology *48*, 389-399.

12. Banerjee,A., Kurbanov,F., Datta,S., Chandra,P.K., Tanaka,Y., Mizokami,M., and Chakravarty,R. (2006). Phylogenetic relatedness and genetic diversity of hepatitis B virus isolates in Eastern India. J. Med. Virol. *78*, 1164-1174.

13. 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.

14. Barlet,V., Zarski,J.P., Thelu,M.A., Bichard,P., and Seigneurin,J.M. (1994). Different prevalence of precore mutants in five members of a hepatitis-B-virus-infected family: evidence for a precore variant type in an asymptomatic anti-HBs patient. J. Hepatol. *21*, 797-805.

15. Bartenschlager,R. and Schaller,H. (1988). The amino-terminal domain of the hepadnaviral P-gene encodes the terminal protein (genome-linked protein) believed to prime reverse transcription. EMBO J. *7*, 4185-4192.

16. Baumert,T.F., Marrone,A., Vergalla,J., and Liang,T.J. (1998). Naturally occurring mutations define a novel function of the hepatitis B virus core promoter in core protein expression. J. Virol. *72*, 6785-6795.

17. Beasley,R.P., Hwang,L.Y., Lin,C.C., and Chien,C.S. (1981). Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22 707 men in Taiwan. Lancet *2*, 1129-1133.

18. Beck,J. and Nassal,M. (2007). Hepatitis B virus replication. World J. Gastroenterol. *13*, 48-64.

19. Beckel-Mitchener,A. and Summers,J. (1997). A novel transcriptional element in circular DNA monomers of the duck hepatitis B virus. J. Virol. *71*, 7917-7922.

20. 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 U. S A *91*, 10350-10354.

21. 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.

22. Bhat,R.A., Ulrich,P.P., and Vyas,G.N. (1990). Molecular characterization of a new variant of hepatitis B virus in a persistently infected homosexual man. Hepatology *11*, 271-276.

23. Bock,C.T., Tillmann,H.L., Maschek,H.J., Manns,M.P., and Trautwein,C. (1997). A preS mutation isolated from a patient with chronic hepatitis B infection leads to virus retention and misassembly. Gastroenterology *113*, 1976-1982.

24. Bonino,F., Rosina,F., Rizzetto,M., Rizzi,R., Chiaberge,E., Tardanico,R., Callea,F., and Verme,G. (1986). Chronic hepatitis in HBsAg carriers with serum HBV-DNA and anti-HBe. Gastroenterology *90*, 1268-1273.

25. Bonvin,M., Achermann,F., Greeve,I., Stroka,D., Keogh,A., Inderbitzin,D., Candinas,D., Sommer,P., Wain-Hobson,S., Vartanian,J.P., and Greeve,J. (2006). Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. Hepatology *43*, 1364-1374.

26. Bosch,V., Bartenschlager,R., Radziwill,G., and Schaller,H. (1988). The duck hepatitis B virus P-gene codes for protein strongly associated with the 5'-end of the viral DNA minus strand. Virology *166*, 475-485.

27. Bowry,T.R. (1983). Seroepidemiology of hepatitis B in an urban population of Nairobi, Kenya. J. Infect. Dis. *148*, 1122.

28. Bowyer,S.M. and Sim,J.G. (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.

29. Bowyer,S.M., van,S.L., Kew,M.C., and Sim,J.G. (1997). A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. J. Gen. Virol. *78 (Pt 7)*, 1719-1729.

30. Bruss,V. (1997). A short linear sequence in the pre-S domain of the large hepatitis B virus envelope protein required for virion formation. J Virol. *71*, 9350-9357.

31. Bruss,V. (2004). Envelopment of the hepatitis B virus nucleocapsid. Virus Res. *106*, 199-209.

32. Bruss,V. and Ganem,D. (1991). The role of envelope proteins in hepatitis B virus assembly. Proc. Natl. Acad. Sci. U. S. A *88*, 1059-1063.

33. Bruss,V. and Gerlich,W.H. (1988). Formation of transmembraneous hepatitis B e-antigen by cotranslational in vitro processing of the viral precore protein. Virology *163*, 268-275.

34. Bruss,V., Hagelstein,J., Gerhardt,E., and Galle,P.R. (1996). Myristylation of the large surface protein is required for hepatitis B virus in vitro infectivity. Virology *218*, 396-399.

35. Buckwold,V.E., Xu,Z., Chen,M., Yen,T.S., and Ou,J.H. (1996). Effects of a naturally occurring mutation in the hepatitis B virus basal core

promoter on precore gene expression and viral replication. J. Virol. *70*, 5845-5851.

36. Buckwold,V.E., Xu,Z., Yen,T.S., and Ou,J.H. (1997). Effects of a frequent double-nucleotide basal core promoter mutation and its putative single-nucleotide precursor mutations on hepatitis B virus gene expression and replication. J. Gen. Virol. *78 (Pt 8)*, 2055-2065.

37. Budkowska,A., Dubreuil,P., Capel,F., and Pillot,J. (1986). Hepatitis B virus pre-S gene-encoded antigenic specificity and anti-pre-S antibody: relationship between anti-pre-S response and recovery. Hepatology *6*, 360-368.

38. Budkowska,A., Dubreuil,P., Poynard,T., Marcellin,P., Loriot,M.A., Maillard,P., and Pillot,J. (1992). Anti-pre-S responses and viral clearance in chronic hepatitis B virus infection. Hepatology *15*, 26-31.

39. Carman,W.F., Jacyna,M.R., Hadziyannis,S., Karayiannis,P., McGarvey,M.J., Makris,A., and Thomas,H.C. (1989). Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. Lancet *2*, 588-591.

40. Carman,W.F., Korula,J., Wallace,L., MacPhee,R., Mimms,L., and Decker,R. (1995). Fulminant reactivation of hepatitis B due to envelope protein mutant that escaped detection by monoclonal HBsAg ELISA. Lancet *345*, 1406-1407.

41. Carman,W.F., Van Deursen,F.J., Mimms,L.T., Hardie,D., Coppola,R., Decker,R., and Sanders,R. (1997). The prevalence of surface antigen variants of hepatitis B virus in Papua New Guinea, South Africa, and Sardinia. Hepatology *26*, 1658-1666.

42. 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. Lancet *336*, 325-329.

43. Caselmann,W.H., Meyer,M., Kekule,A.S., Lauer,U., Hofschneider,P.H., and Koshy,R. (1990). A trans-activator function is generated by integration of hepatitis B virus preS/S sequences in human hepatocellular carcinoma DNA. Proc. Natl. Acad. Sci. U. S. A *87*, 2970-2974.

44. Cattaneo,R., Will,H., and Schaller,H. (1984). Hepatitis B virus transcription in the infected liver. EMBO J. *3*, 2191-2196.

45. Chan,H.L., Tsui,S.K., Tse,C.H., Ng,E.Y., Au,T.C., Yuen,L., Bartholomeusz,A., Leung,K.S., Lee,K.H., Locarnini,S., and Sung,J.J. (2005). Epidemiological and virological characteristics of 2 subgroups of hepatitis B virus genotype C. J. Infect. Dis. *191*, 2022-2032.

46. Chang,J.J. and Lewin,S.R. (2007). Immunopathogenesis of hepatitis B virus infection. Immunol. Cell Biol. *85*, 16-23.

47. Chang,L.J., Ganem,D., and Varmus,H.E. (1990). Mechanism of translation of the hepadnaviral polymerase (P) gene. Proc. Natl. Acad. Sci. U. S. A *87*, 5158-5162.

48. Chang,M.H., Chen,C.J., Lai,M.S., Hsu,H.M., Wu,T.C., Kong,M.S., Liang,D.C., Shau,W.Y., and Chen,D.S. (1997). Universal hepatitis B vaccination in Taiwan and the incidence of hepatocellular carcinoma in children. Taiwan Childhood Hepatoma Study Group. N. Engl. J. Med. *336*, 1855-1859.

49. Chen,C.Y., Crowther,C., Kew,M.C., and Kramvis,A. (2008). A valine to phenylalanine mutation in the precore region of hepatitis B virus causes intracellular retention and impaired secretion of HBe-antigen 1. Hepatol. Res.

50. 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.

51. Chen,I.H., Huang,C.J., and Ting,L.P. (1995). Overlapping initiator and TATA box functions in the basal core promoter of hepatitis B virus. J. Virol. *69*, 3647-3657.

52. Chen,M., Sallberg,M., Hughes,J., Jones,J., Guidotti,L.G., Chisari,F.V., Billaud,J.N., and Milich,D.R. (2005). Immune tolerance split between hepatitis B virus precore and core proteins. J. Virol. *79*, 3016-3027.

53. Chen,Y. and Marion,P.L. (1996). Amino acids essential for RNase H activity of hepadnaviruses are also required for efficient elongation of minus-strand viral DNA. J. Virol. *70*, 6151-6156.

54. Cheng,A.S., Chan,H.L., To,K.F., Leung,W.K., Chan,K.K., Liew,C.T., and Sung,J.J. (2004). Cyclooxygenase-2 pathway correlates with vascular endothelial growth factor expression and tumor angiogenesis in hepatitis B virus-associated hepatocellular carcinoma. Int. J. Oncol. *24*, 853-860.

55. Chisari,F.V. and Ferrari,C. (1995). Hepatitis B virus immunopathology. Springer Semin. Immunopathol. *17*, 261-281.

56. 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.

57. Coursaget,P., Adamowicz,P., Bourdil,C., Yvonnet,B., Buisson,Y., Barres,J.L., Saliou,P., Chiron,J.P., and Mar,I.D. (1988). Anti-pre-S2 antibodies in natural hepatitis B virus infection and after immunization. Vaccine *6*, 357-361.

58. Crowther,R.A., Kiselev,N.A., Bottcher,B., Berriman,J.A., Borisova,G.P., Ose,V., and Pumpens,P. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell *77*, 943-950.

59. Cui,C., Shi,J., Hui,L., Xi,H., Zhuoma, Quni, Tsedan, and Hu,G. (2002). The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. J. Gen. Virol. *83*, 2773-2777.

60. Dane,D.S., Cameron,C.H., and Briggs,M. (1970). Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. Lancet *1*, 695-698.

61. De,F.S., Ruvoletto,M.G., Verdoliva,A., Ruvo,M., Raucci,A., Marino,M., Senatore,S., Cassani,G., Alberti,A., Pontisso,P., and Fassina,G. (2001). Cloning and expression of a novel hepatitis B virus-binding protein from HepG2 cells. J. Biol. Chem. *276*, 36613-36623.

62. Dejean,A., Sonigo,P., Wain-Hobson,S., and Tiollais,P. (1984). Specific hepatitis B virus integration in hepatocellular carcinoma DNA through a viral 11-base-pair direct repeat
2. Proc. Natl. Acad. Sci U. S A *81*, 5350-5354.

63. Devesa,M., Rodriguez,C., Leon,G., Liprandi,F., and Pujol,F.H. (2004). Clade analysis and surface antigen polymorphism of hepatitis B virus American genotypes. J. Med. Virol. *72*, 377-384.

64. Dienstag,J.L., Perrillo,R.P., Schiff,E.R., Bartholomew,M., Vicary,C., and Rubin,M. (1995). A preliminary trial of lamivudine for chronic hepatitis B infection. N. Engl. J. Med. *333*, 1657-1661.

65. Dieterich,D.T. (2007). Special considerations and treatment of patients with HBV-HIV coinfection. Antivir. Ther. *12 Suppl 3*, H43-H51.

66. DiNardo,S., Voelkel,K., and Sternglanz,R. (1984). DNA topoisomerase II mutant of Saccharomyces cerevisiae: topoisomerase II is required for segregation of daughter molecules at the termination of DNA replication
5. Proc. Natl. Acad. Sci U. S A *81*, 2616-2620.

67. Du,H., Li,T., Zhang,H.Y., He,Z.P., Dong,Q.M., Duan,X.Z., and Zhuang,H. (2007). Correlation of hepatitis B virus (HBV) genotypes and mutations in basal core promoter/precore with clinical features of chronic HBV infection. Liver Int. *27*, 240-246.

68. el-Deiry,W.S., Harper,J.W., O'Connor,P.M., Velculescu,V.E., Canman,C.E., Jackman,J., Pietenpol,J.A., Burrell,M., Hill,D.E., Wang,Y., and . (1994). WAF1/CIP1 is induced in p53-mediated G1 arrest and apoptosis. Cancer Res. *54*, 1169-1174.

69. Elmore,L.W., Hancock,A.R., Chang,S.F., Wang,X.W., Chang,S., Callahan,C.P., Geller,D.A., Will,H., and Harris,C.C. (1997). Hepatitis B

virus X protein and p53 tumor suppressor interactions in the modulation of apoptosis. Proc. Natl. Acad. Sci U. S A *94*, 14707-14712.

70. Fan,Y.F., Lu,C.C., Chang,Y.C., Chang,T.T., Lin,P.W., Lei,H.Y., and Su,I.J. (2000). Identification of a pre-S2 mutant in hepatocytes expressing a novel marginal pattern of surface antigen in advanced diseases of chronic hepatitis B virus infection. J. Gastroenterol. Hepatol. *15*, 519-528.

71. Fan,Y.F., Lu,C.C., Chen,W.C., Yao,W.J., Wang,H.C., Chang,T.T., Lei,H.Y., Shiau,A.L., and Su,I.J. (2001). Prevalence and significance of hepatitis B virus (HBV) pre-S mutants in serum and liver at different replicative stages of chronic HBV infection. Hepatology *33*, 277-286.

72. Fares,M.A. and Holmes,E.C. (2002). A revised evolutionary history of hepatitis B virus (HBV). J Mol. Evol. *54*, 807-814.

73. Feitelson,M.A. (1994). Biology of hepatitis B virus variants. Lab Invest *71*, 324-349.

74. Felsenstein, J. PHYLIP (Phylogeny Inference Package). 1995. Department of Genetics, University of Washington, Seattle, Washington.
Ref Type: Computer Program

75. Ferrari,C., Bertoletti,A., Penna,A., Cavalli,A., Valli,A., Missale,G., Pilli,M., Fowler,P., Giuberti,T., Chisari,F.V., and . (1991). Identification of immunodominant T cell epitopes of the hepatitis B virus nucleocapsid antigen. J Clin. Invest *88*, 214-222.

76. Ferrer-Orta,C., Arias,A., Escarmis,C., and Verdaguer,N. (2006). A comparison of viral RNA-dependent RNA polymerases. Curr. Opin. Struct. Biol. *16*, 27-34.

77. Francis,D.P., Favero,M.S., and Maynard,J.E. (1981). Transmission of hepatitis B virus. Semin. Liver Dis. *1*, 27-32.

78. Fujiwara,K., Tanaka,Y., Paulon,E., Orito,E., Sugiyama,M., Ito,K., Ueda,R., Mizokami,M., and Naoumov,N.V. (2005). Novel type of hepatitis B virus mutation: replacement mutation involving a hepatocyte nuclear factor 1 binding site tandem repeat in chronic hepatitis B virus genotype E. J. Virol. *79*, 14404-14410.

79. 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.

80. Gallina,A., Bonelli,F., Zentilin,L., Rindi,G., Muttini,M., and Milanesi,G. (1989). A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. J Virol. *63*, 4645-4652.

81. Ganem,D. and Varmus,H.E. (1987). The molecular biology of the hepatitis B viruses. Annu. Rev. Biochem. *56*, 651-693.

82. Garcia,A.D., Ostapchuk,P., and Hearing,P. (1993). Functional interaction of nuclear factors EF-C, HNF-4, and RXR alpha with hepatitis B virus enhancer I. J. Virol. *67*, 3940-3950.

83. Georgi-Geisberger,P., Berns,H., Loncarevic,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.

84. Gerber,M.A. and Thung,S.N. (1985). Molecular and cellular pathology of hepatitis B. Lab Invest *52*, 572-590.

85. Gerner,P., Lausch,E., Friedt,M., Tratzmuller,R., Spangenberg,C., and Wirth,S. (1999). Hepatitis B virus core promoter mutations in children with multiple anti-HBe/HBeAg reactivations result in enhanced promoter activity. J. Med. Virol. *59*, 415-423.

86. Gerner,P., Schafer,H.M., Prange,R., Pravitt,D., and Wirth,S. (2003). Functional analysis of a rare HBV deletion mutant in chronically infected children. Pediatr. Res. *53*, 891-897.

87. Gilles,P.N., Guerrette,D.L., Ulevitch,R.J., Schreiber,R.D., and Chisari,F.V. (1992). HBsAg retention sensitizes the hepatocyte to injury by physiological concentrations of interferon-gamma
1. Hepatology *16*, 655-663.

88. Glebe,D. and Urban,S. (2007). Viral and cellular determinants involved in hepadnaviral entry. World J. Gastroenterol. *13*, 22-38.

89. Graves, B. GraphPad. [5]. 2003. GraphPad Software.Inc San Diego. Ref Type: Computer Program

90. Guarnieri,M., Kim,K.H., Bang,G., Li,J., Zhou,Y., Tang,X., Wands,J., and Tong,S. (2006). Point mutations upstream of hepatitis B virus core gene affect DNA replication at the step of core protein expression. J. Virol. *80*, 587-595.

91. Guidotti,L.G., Ishikawa,T., Hobbs,M.V., Matzke,B., Schreiber,R., and Chisari,F.V. (1996a). Intracellular inactivation of the hepatitis B virus by cytotoxic T lymphocytes
5. Immunity. *4*, 25-36.

92. Guidotti,L.G., Matzke,B., Pasquinelli,C., Shoenberger,J.M., Rogler,C.E., and Chisari,F.V. (1996b). The hepatitis B virus (HBV) precore protein inhibits HBV replication in transgenic mice. J. Virol. *70*, 7056-7061.

93. Gunther,S. (2006). Genetic variation in HBV infection: genotypes and mutants. J. Clin. Virol. *36 Suppl 1*, S3-S11.

94. Gunther,S., Baginski,S., Kissel,H., Reinke,P., Kruger,D.H., Will,H., and Meisel,H. (1996a). Accumulation and persistence of hepatitis B virus core gene deletion mutants in renal transplant patients are associated with end-stage liver disease. Hepatology *24*, 751-758.

95. Gunther,S., Fischer,L., Pult,I., Sterneck,M., and Will,H. (1999). Naturally occurring variants of hepatitis B virus. Adv. Virus Res. *52*, 25-137.

96. 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.

97. Gunther,S., Piwon,N., Iwanska,A., Schilling,R., Meisel,H., and Will,H. (1996b). Type, prevalence, and significance of core promoter/enhancer II mutations in hepatitis B viruses from immunosuppressed patients with severe liver disease. J. Virol. *70*, 8318-8331.

98. Gunther,S., Piwon,N., Jung,A., Iwanska,A., Schmitz,H., and Will,H. (2000). Enhanced replication contributes to enrichment of hepatitis B virus with a deletion in the core gene. Virology *273*, 286-299.

99. Gunther,S., Sommer,G., Iwanska,A., and Will,H. (1997a). Heterogeneity and common features of defective hepatitis B virus genomes derived from spliced pregenomic RNA. Virology *238*, 363-371.

100. Gunther,S., Sommer,G., Plikat,U., Iwanska,A., Wain-Hobson,S., Will,H., and Meyerhans,A. (1997b). Naturally occurring hepatitis B virus genomes bearing the hallmarks of retroviral G-->A hypermutation. Virology *235*, 104-108.

101. Guo,W., Chen,M., Yen,T.S., and Ou,J.H. (1993). Hepatocyte-specific expression of the hepatitis B virus core promoter depends on both positive and negative regulation. Mol. Cell Biol. *13*, 443-448.

102. Hall, T. BioEdit. [7.0.9]. 2007. A subsidiary of Isis Pharmaceuticals 1891 Rutherford Road Carlsbad, CA 92008 760-603-2375, Ibis Biosciences.
Ref Type: Computer Program

103. Hampton,R.Y. (2000). ER stress response: getting the UPR hand on misfolded proteins. Curr. Biol. *10*, R518-R521.

104. Han,J.A., Kim,J.I., Ongusaha,P.P., Hwang,D.H., Ballou,L.R., Mahale,A., Aaronson,S.A., and Lee,S.W. (2002). P53-mediated induction of Cox-2 counteracts p53- or genotoxic stress-induced apoptosis
6. EMBO J *21*, 5635-5644.

105. Hantz,O., Baginski,I., Fourel,I., Chemin,I., and Trepo,C. (1992). Viral spliced RNA are produced, encapsidated and reverse transcribed during in vivo woodchuck hepatitis virus infection. Virology *190*, 193-200.

106. Henkler,F., Hoare,J., Waseem,N., Goldin,R.D., McGarvey,M.J., Koshy,R., and King,I.A. (2001). Intracellular localization of the hepatitis B virus HBx protein. J. Gen. Virol. *82*, 871-882.

107. Hildt,E., Munz,B., Saher,G., Reifenberg,K., and Hofschneider,P.H. (2002). The PreS2 activator MHBs(t) of hepatitis B virus activates c-raf-1/Erk2 signaling in transgenic mice. EMBO J. *21*, 525-535.

108. Hildt,E., Saher,G., Bruss,V., and Hofschneider,P.H. (1996a). The hepatitis B virus large surface protein (LHBs) is a transcriptional activator. Virology *225*, 235-239.

109. Hildt,E., Urban,S., Eckerskorn,C., and Hofschneider,P.H. (1996b). Isolation of highly purified, functional carboxy-terminally truncated hepatitis B virus middle surface protein activators from eucaryotic expression systems. Hepatology *24*, 502-507.

110. Hildt,E., Urban,S., and Hofschneider,P.H. (1995). Characterization of essential domains for the functionality of the MHBst transcriptional activator and identification of a minimal MHBst activator. Oncogene *11*, 2055-2066.

111. Hildt,E., Urban,S., Lauer,U., Hofschneider,P.H., and Kekule,A.S. (1993). ER-localization and functional expression of the HBV transactivator MHBst. Oncogene *8*, 3359-3367.

112. Hino,O., Ohtake,K., and Rogler,C.E. (1989). Features of two hepatitis B virus (HBV) DNA integrations suggest mechanisms of HBV integration
4. J Virol. *63*, 2638-2643.

113. Hino,O., Shows,T.B., and Rogler,C.E. (1986). Hepatitis B virus integration site in hepatocellular carcinoma at chromosome 17;18 translocation
2. Proc. Natl. Acad. Sci U. S A *83*, 8338-8342.

114. Hoare,J., Henkler,F., Dowling,J.J., Errington,W., Goldin,R.D., Fish,D., and McGarvey,M.J. (2001). Subcellular localisation of the X protein in HBV infected hepatocytes. J. Med. Virol. *64*, 419-426.

115. Hollinger,F.B., Goyal,R.K., Hersh,T., Powell,H.C., Schulman,R.J., and Malnick,J.L. (1972). Immune response to hepaitis virus type B in Down's syndrome and other mentally retarded patients. Am. J Epidemiol. *95*, 356-362.

116. Honigwachs,J., Faktor,O., Dikstein,R., Shaul,Y., and Laub,O. (1989). Liver-specific expression of hepatitis B virus is determined by the

combined action of the core gene promoter and the enhancer. J. Virol. *63*, 919-924.

117. Horikita,M., Itoh,S., Yamamoto,K., Shibayama,T., Tsuda,F., and Okamoto,H. (1994). Differences in the entire nucleotide sequence between hepatitis B virus genomes from carriers positive for antibody to hepatitis B e antigen with and without active disease. J. Med. Virol. *44*, 96-103.

118. Hou,J., Karayiannis,P., Waters,J., Luo,K., Liang,C., and Thomas,H.C. (1995). A unique insertion in the S gene of surface antigen--negative hepatitis B virus Chinese carriers. Hepatology *21*, 273-278.

119. Hou,J., Lin,Y., Waters,J., Wang,Z., Min,J., Liao,H., Jiang,J., Chen,J., Luo,K., and Karayiannis,P. (2002). Detection and significance of a G1862T variant of hepatitis B virus in Chinese patients with fulminant hepatitis. J Gen. Virol. *83*, 2291-2298.

120. hppt and www.stanford.edu/group/virus/1999/tchang/replication.htm. hppt//www.stanford.edu/group/virus/1999/tchang/replication.htm. Human Virology at Stanford . 19-12-2007.
Ref Type: Electronic Citation

121. Hsia,C.C., Yuwen,H., and Tabor,E. (1996). Hot-spot mutations in hepatitis B virus X gene in hepatocellular carcinoma. Lancet *348*, 625-626.

122. Hsieh,Y.H., Su,I.J., Wang,H.C., Chang,W.W., Lei,H.Y., Lai,M.D., Chang,W.T., and Huang,W. (2004c). Pre-S mutant surface antigens in chronic hepatitis B virus infection induce oxidative stress and DNA damage. Carcinogenesis *25*, 2023-2032.

123. Hsieh,Y.H., Su,I.J., Wang,H.C., Chang,W.W., Lei,H.Y., Lai,M.D., Chang,W.T., and Huang,W. (2004b). Pre-S mutant surface antigens in chronic hepatitis B virus infection induce oxidative stress and DNA damage. Carcinogenesis *25*, 2023-2032.

124. Hsieh,Y.H., Su,I.J., Wang,H.C., Chang,W.W., Lei,H.Y., Lai,M.D., Chang,W.T., and Huang,W. (2004a). Pre-S mutant surface antigens in chronic hepatitis B virus infection induce oxidative stress and DNA damage. Carcinogenesis *25*, 2023-2032.

125. Hsieh,Y.H., Su,I.J., Wang,H.C., Tsai,J.H., Huang,Y.J., Chang,W.W., Lai,M.D., Lei,H.Y., and Huang,W. (2007). Hepatitis B virus pre-S2 mutant surface antigen induces degradation of cyclin-dependent kinase inhibitor p27Kip1 through c-Jun activation domain-binding protein 1. Mol. Cancer Res. *5*, 1063-1072.

126. Huang,H.L., Jeng,K.S., Hu,C.P., Tsai,C.H., Lo,S.J., and Chang,C. (2000). Identification and characterization of a structural protein of

hepatitis B virus: a polymerase and surface fusion protein encoded by a spliced RNA. Virology *275*, 398-410.

127. Huang,J. and Liang,T.J. (1993). A novel hepatitis B virus (HBV) genetic element with Rev response element-like properties that is essential for expression of HBV gene products
2. Mol. Cell Biol. *13*, 7476-7486.

128. 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.

129. Huang,Z.M. and Yen,T.S. (1993). Dysregulated surface gene expression from disrupted hepatitis B virus genomes. J. Virol. *67*, 7032-7040.

130. Huang,Z.M. and Yen,T.S. (1995). Role of the hepatitis B virus posttranscriptional regulatory element in export of intronless transcripts. Mol. Cell Biol. *15*, 3864-3869.

131. Hung,J.H., Su,I.J., Lei,H.Y., Wang,H.C., Lin,W.C., Chang,W.T., Huang,W., Chang,W.C., Chang,Y.S., Chen,C.C., and Lai,M.D. (2004a). Endoplasmic reticulum stress stimulates the expression of cyclooxygenase-2 through activation of NF-kappaB and pp38 mitogen-activated protein kinase. J. Biol. Chem. *279*, 46384-46392.

132. Hung,J.H., Su,I.J., Lei,H.Y., Wang,H.C., Lin,W.C., Chang,W.T., Huang,W., Chang,W.C., Chang,Y.S., Chen,C.C., and Lai,M.D. (2004b). Endoplasmic reticulum stress stimulates the expression of cyclooxygenase-2 through activation of NF-kappaB and pp38 mitogen-activated protein kinase. J. Biol. Chem. *279*, 46384-46392.

133. Huy,T.T., Ushijima,H., Quang,V.X., Win,K.M., Luengrojanakul,P., Kikuchi,K., Sata,T., and Abe,K. (2004). Genotype C of hepatitis B virus can be classified into at least two subgroups. J. Gen. Virol. *85*, 283-292.

134. Huy,T.T., Ushijima,H., Sata,T., and Abe,K. (2006). Genomic characterization of HBV genotype F in Bolivia: genotype F subgenotypes correlate with geographic distribution and T(1858) variant. Arch. Virol. *151*, 589-597.

135. Huy,T.T., Ushijima,H., Win,K.M., Luengrojanakul,P., Shrestha,P.K., Zhong,Z.H., Smirnov,A.V., Taltavull,T.C., Sata,T., and Abe,K. (2003). High prevalence of hepatitis B virus pre-s mutant in countries where it is endemic and its relationship with genotype and chronicity. J. Clin. Microbiol. *41*, 5449-5455.

136. Hwang,W.L. and Su,T.S. (1998). Translational regulation of hepatitis B virus polymerase gene by termination-reinitiation of an upstream

minicistron in a length-dependent manner. J. Gen. Virol. *79 (Pt 9)*, 2181-2189.

137. Jackson,I.J. (1991). A reappraisal of non-consensus mRNA splice sites 2. Nucleic Acids Res. *19*, 3795-3798.

138. Jin,Y.H., Yoo,K.J., Lee,Y.H., and Lee,S.K. (2000). Caspase 3-mediated cleavage of p21WAF1/CIP1 associated with the cyclin A-cyclin-dependent kinase 2 complex is a prerequisite for apoptosis in SK-HEP-1 cells
5. J Biol. Chem. *275*, 30256-30263.

139. Junker-Niepmann,M., Bartenschlager,R., and Schaller,H. (1990). A short cis-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. EMBO J. *9*, 3389-3396.

140. Kajiya,Y., Hamasaki,K., Nakata,K., Nakagawa,Y., Miyazoe,S., Takeda,Y., Ohkubo,K., Ichikawa,T., Nakao,K., Kato,Y., and Eguchi,K. (2002). Full-length sequence and functional analysis of hepatitis B virus genome in a virus carrier: a case report suggesting the impact of pre-S and core promoter mutations on the progression of the disease. J. Viral Hepat. *9*, 149-156.

141. 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
14. Gastroenterology *118*, 554-559.

142. Kato,H., Fujiwara,K., Gish,R.G., Sakugawa,H., Yoshizawa,H., Sugauchi,F., Orito,E., Ueda,R., Tanaka,Y., Kato,T., Miyakawa,Y., and Mizokami,M. (2005). Classifying genotype F of hepatitis B virus into F1 and F2 subtypes. World J. Gastroenterol. *11*, 6295-6304.

143. Kato,H., Orito,E., Gish,R.G., Bzowej,N., Newsom,M., Sugauchi,F., Suzuki,S., Ueda,R., Miyakawa,Y., and Mizokami,M. (2002). Hepatitis B e antigen in sera from individuals infected with hepatitis B virus of genotype G. Hepatology *35*, 922-929.

144. Kekule,A.S., Lauer,U., Meyer,M., Caselmann,W.H., Hofschneider,P.H., and Koshy,R. (1990). The preS2/S region of integrated hepatitis B virus DNA encodes a transcriptional transactivator. Nature *343*, 457-461.

145. Kekule,A.S., Lauer,U., Weiss,L., Luber,B., and Hofschneider,P.H. (1993). Hepatitis B virus transactivator HBx uses a tumour promoter signalling pathway. Nature *361*, 742-745.

146. Kew,M.C. (1981). The hepatitis-B virus and hepatocellular carcinoma. Semin. Liver Dis. *1*, 59-67.

147. Kew,M.C. (1989). Hepatocellular carcinoma with and without cirrhosis. A comparison in southern African blacks. Gastroenterology *97*, 136-139.

148. Kew,M.C. (1992). Chronic hepatitis B virus infection and hepatocellular carcinoma in Africa. S Afr J Sci *88*, -524.

149. Kew,M.C. (2002). Hepatitis B virus as a cause of hepatocellular carcinoma, with particular reference to sub-Saharan Africa. S Afr J Epidemiol Infect *17*, 34-39.

150. Kew,M.C. (2006). Epidemiology of HBV infection and HBV-related hepatocellular carcinoma in Africa: natural history and clinical outcome. Science Series *1*, 84-88.

151. Kew,M.C., Desmyter,J., Bradburne,A.F., and Macnab,G.M. (1979). Hepatitis B virus infection in southern African blacks with hepatocellular cancer. J. Natl. Cancer Inst. *62*, 517-520.

152. Kew,M.C., Kassianides,C., Berger,E.L., Song,E., and Dusheiko,G.M. (1987). Prevalence of chronic hepatitis B virus infection in pregnant black women living in Soweto. J. Med. Virol. *22*, 263-268.

153. Kew,M.C., Kramvis,A., Yu,M.C., Arakawa,K., and Hodkinson,J. (2005). Increased hepatocarcinogenic potential of hepatitis B virus genotype A in Bantu-speaking sub-saharan Africans. J. Med. Virol. *75*, 513-521.

154. 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. U. S. A *90*, 10211-10215.

155. Kew,M.C. and Paterson,A.C. (1985). Unusual clinical presentations of hepatocellular carcinoma. Trop. Gastroenterol. *6*, 10-22.

156. Kew,M.C., Ray,M.B., Desmet,V.J., and Desmyter,J. (1980). Hepatitis-B surface antigen in tumour tissue and non-tumorous liver in black patients with hepatocellular carcinoma. Br. J. Cancer *41*, 399-406.

157. 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.

158. Kim,H., Jee,Y., Mun,H.S., Park,J.H., Yoon,J.H., Kim,Y.J., Lee,H.S., Hyun,J.W., Hwang,E.S., Cha,C.Y., Kook,Y.H., and Kim,B.J. (2007). Characterization of Two Hepatitis B Virus Populations in a Single Korean Hepatocellular Carcinoma Patient with an HBeAg-Negative Serostatus: A Novel X-Gene-Deleted Strain with Inverted Duplication Sequences of Upstream Enhancer Site II. Intervirology *50*, 273-280.

159. Kimbi,G.C., Kramvis,A., and Kew,M.C. (2004). Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa. J. Gen. Virol. *85*, 1211-1220.

160. Kimbi, GC. Full genome analysis and functional characterization of mutants of hepatitis B virus isolates fron southern African Blacks. 2005.
Ref Type: Thesis/Dissertation

161. Kirk,G.D., Lesi,O.A., Mendy,M., Akano,A.O., Sam,O., Goedert,J.J., Hainaut,P., Hall,A.J., Whittle,H., and Montesano,R. (2004). The Gambia Liver Cancer Study: Infection with hepatitis B and C and the risk of hepatocellular carcinoma in West Africa. Hepatology *39*, 211-219.

162. Knowles,B.B., Searls,D.B., and Aden,D.P. (1984). Human hepatoma derived cells. In Advances in Hepatitis Research (Chisari, F.V., ed), Masson Publication, New York).

163. Koike,K., Moriya,K., Yotsuyanagi,H., Shintani,Y., Fujie,H., Tsutsumi,T., and Kimura,S. (1998). Compensatory apoptosis in preneoplastic liver of a transgenic mouse model for viral hepatocarcinogenesis. Cancer Lett. *134*, 181-186.

164. Konopka,A.K. (1988). Compilation of DNA strand exchange sites for non-homologous recombination in somatic cells. Nucleic Acids Res. *16*, 1739-1758.

165. Kozak,M. (1986). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. U. S. A *83*, 2850-2854.

166. Kramvis,A., Arakawa,K., Yu,M.C., Nogueira,R., Stram,D.O., and Kew,M.C. (2008). Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus
2. J Med. Virol. *80*, 27-46.

167. 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.

168. Kramvis,A., Kew,M., and Francois,G. (2005a). Hepatitis B virus genotypes. Vaccine *23*, 2409-2423.

169. Kramvis,A. and Kew,M.C. (1998). Structure and function of the encapsidation signal of hepadnaviridae. J. Viral Hepat. *5*, 357-367.

170. Kramvis,A. and Kew,M.C. (1999). The core promoter of hepatitis B virus. J. Viral Hepat. *6*, 415-427.

171. Kramvis,A. and Kew,M.C. (2005). Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. J. Viral Hepat. *12*, 456-464.

172. Kramvis,A. and Kew,M.C. (2007a). Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes. Hepatol. Res. *37*, S9-S19.

173. Kramvis,A. and Kew,M.C. (2007b). Molecular characterization of subgenotype A1 (subgroup Aa) of hepatitis B virus. Hepatol. Res. *37*, S27-S32.

174. Kramvis,A., Kew,M.C., and Bukofzer,S. (1998). Hepatitis B virus precore mutants in serum and liver of Southern African Blacks with hepatocellular carcinoma. J. Hepatol. *28*, 132-141.

175. Kramvis,A., Restorp,K., Norder,H., Botha,J.F., Magnius,L.O., and Kew,M.C. (2005b). Full genome analysis of hepatitis B virus genotype E strains from South-Western Africa and Madagascar reveals low genetic variability. J. Med. Virol. *77*, 47-52.

176. Kramvis,A., Weitzmann,L., Owiredu,W.K., and Kew,M.C. (2002). Analysis of the complete genome of subgroup A' hepatitis B virus isolates from South Africa. J. Gen. Virol. *83*, 835-839.

177. Kuang,S.Y., Jackson,P.E., Wang,J.B., Lu,P.X., Munoz,A., Qian,G.S., Kensler,T.W., and Groopman,J.D. (2004). Specific mutations of hepatitis B virus in plasma predict liver cancer development. Proc. Natl. Acad. Sci. U. S. A *101*, 3575-3580.

178. Kurbanov,F., Tanaka,Y., Fujiwara,K., Sugauchi,F., Mbanya,D., Zekeng,L., Ndembi,N., Ngansop,C., Kaptue,L., Miura,T., Ido,E., Hayami,M., Ichimura,H., and Mizokami,M. (2005). A new subtype (subgenotype) Ac (A3) of hepatitis B virus and recombination between genotypes A and E in Cameroon. J. Gen. Virol. *86*, 2047-2056.

179. Kuroki,K., Cheung,R., Marion,P.L., and Ganem,D. (1994). A cell surface protein that binds avian hepatitis B virus particles. J. Virol. *68*, 2091-2096.

180. Kuroki,K., Eng,F., Ishikawa,T., Turck,C., Harada,F., and Ganem,D. (1995). gp180, a host cell glycoprotein that binds duck hepatitis B virus particles, is encoded by a member of the carboxypeptidase gene family. J Biol. Chem. *270*, 15022-15028.

181. Kurosaki,M., Enomoto,N., Asahina,Y., Sakuma,I., Ikeda,T., Tozuka,S., Izumi,N., Marumo,F., and Sato,C. (1996). Mutations in the core promoter region of hepatitis B virus in patients with chronic hepatitis B. J. Med. Virol. *49*, 115-123.

182. Lamberts,C., Nassal,M., Velhagen,I., Zentgraf,H., and Schroder,C.H. (1993). Precore-mediated inhibition of hepatitis B virus progeny DNA synthesis. J. Virol. *67*, 3756-3762.

183. Lanford,R.E., Chavez,D., Brasky,K.M., Burns,R.B., III, and Rico-Hesse,R. (1998). Isolation of a hepadnavirus from the woolly monkey, a New World primate. Proc. Natl. Acad. Sci. U. S. A *95*, 5757-5761.

184. Laskus,T., Rakela,J., and Persing,D.H. (1994a). The stem-loop structure of the cis-encapsidation signal is highly conserved in naturally occurring hepatitis B virus variants. Virology *200*, 809-812.

185. Laskus,T., Rakela,J., Tong,M.J., Nowicki,M.J., Mosley,J.W., and Persing,D.H. (1994b). Naturally occurring hepatitis B virus mutants with deletions in the core promoter region. J. Hepatol. *20*, 837-841.

186. Lauer,U., Weiss,L., Hofschneider,P.H., and Kekule,A.S. (1992). The hepatitis B virus pre-S/S(t) transactivator is generated by 3' truncations within a defined region of the S gene. J. Virol. *66*, 5284-5289.

187. Le Bouvier,G.L. (1971). The heterogeneity of Australia antigen. J Infect. Dis. *123*, 671-675.

188. Le Seyec,J., Chouteau,P., Cannie,I., Guguen-Guillouzo,C., and Gripon,P. (1998). Role of the pre-S2 domain of the large envelope protein in hepatitis B virus assembly and infectivity. J. Virol. *72*, 5573-5578.

189. Le Seyec,J., Chouteau,P., Cannie,I., Guguen-Guillouzo,C., and Gripon,P. (1999). Infection process of the hepatitis B virus depends on the presence of a defined sequence in the pre-S1 domain. J Virol. *73*, 2052-2057.

190. Lee,T.K., Man,K., Poon,R.T., Lo,C.M., Ng,I.O., and Fan,S.T. (2004). Disruption of p53-p21/WAF1 cell cycle pathway contributes to progression and worse clinical outcome of hepatocellular carcinoma. Oncol. Rep. *12*, 25-31.

191. Lee,Y.H. and Yun,Y. (1998). HBx protein of hepatitis B virus activates Jak1-STAT signaling. J Biol. Chem. *273*, 25510-25515.

192. Levine,A.J. (1997). p53, the cellular gatekeeper for growth and division 9. Cell *88*, 323-331.

193. Li,J., Buckwold,V.E., Hon,M.W., and Ou,J.H. (1999). Mechanism of suppression of hepatitis B virus precore RNA transcription by a frequent double mutation. J. Virol. *73*, 1239-1244.

194. Li,M., Xie,Y., Wu,X., Kong,Y., and Wang,Y. (1995). HNF3 binds and activates the second enhancer, ENII, of hepatitis B virus. Virology *214*, 371-378.

195. Lian,M., Zhou,X., Chen,B., Li,C., Gu,X., Luo,M., and Zheng,X. (2008). Identification of the critical regions in hepatitis B virus preS required for its stability
1. J Pept. Sci *14*, 307-312.

196. Lin,X., Qian,G.S., Lu,P.X., Wu,L., and Wen,Y.M. (2001). Full-length genomic analysis of hepatitis B virus isolates in a patient progressing from hepatitis to hepatocellular carcinoma. J. Med. Virol. *64*, 299-304.

197. Lindh,M., Hannoun,C., Dhillon,A.P., Norkrans,G., and Horal,P. (1999). Core promoter mutations and genotypes in relation to viral replication and liver damage in East Asian hepatitis B virus carriers. J. Infect. Dis. *179*, 775-782.

198. Liu,M., Chan,C.W., McGilvray,I., Ning,Q., and Levy,G.A. (2001). Fulminant viral hepatitis: molecular and cellular basis, and clinical implications. Expert. Rev. Mol. Med. *2001*, 1-19.

199. Lo,W.Y. and Ting,L.P. (1994). Repression of enhancer II activity by a negative regulatory element in the hepatitis B virus genome. J. Virol. *68*, 1758-1764.

200. Loffler-Mary,H., Werr,M., and Prange,R. (1997). Sequence-specific repression of cotranslational translocation of the hepatitis B virus envelope proteins coincides with binding of heat shock protein Hsc70. Virology *235*, 144-152.

201. Lok,A.S., Akarca,U., and Greene,S. (1994). Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. Proc. Natl. Acad. Sci. U. S. A *91*, 4077-4081.

202. Lopez-Cabrera,M., Letovsky,J., Hu,K.Q., and Siddiqui,A. (1990). Multiple liver-specific factors bind to the hepatitis B virus core/pregenomic promoter: trans-activation and repression by CCAAT/enhancer binding protein. Proc. Natl. Acad. Sci. U. S. A *87*, 5069-5073.

203. Lopez-Cabrera,M., Letovsky,J., Hu,K.Q., and Siddiqui,A. (1991). Transcriptional factor C/EBP binds to and transactivates the enhancer element II of the hepatitis B virus. Virology *183*, 825-829.

204. Lu,C.C., Chen,M., Ou,J.H., and Yen,T.S. (1995). Key role of a CCAAT element in regulating hepatitis B virus surface protein expression. Virology *206*, 1155-1158.

205. 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.

206. Macnab,G.M., Alexander,J.J., Lecatsas,G., Bey,E.M., and Urbanowicz,J.M. (1976). Hepatitis B surface antigen produced by a human hepatoma cell line. Br. J. Cancer *34*, 509-515.

207. Magnius,L.O. and Norder,H. (1995). Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. Intervirology *38*, 24-34.

208. Maini,M.K., Boni,C., Lee,C.K., Larrubia,J.R., Reignat,S., Ogg,G.S., King,A.S., Herberg,J., Gilson,R., Alisa,A., Williams,R., Vergani,D., Naoumov,N.V., Ferrari,C., and Bertoletti,A. (2000). The role of virus-specific CD8(+) cells in liver damage and viral control during persistent hepatitis B virus infection. J Exp. Med. *191*, 1269-1280.

209. Makuwa,M., Souquiere,S., Telfer,P., Apetrei,C., Vray,M., Bedjabaga,I., Mouinga-Ondeme,A., Onanga,R., Marx,P.A., Kazanji,M., Roques,P., and Simon,F. (2006). Identification of hepatitis B virus subgenotype A3 in rural Gabon. J. Med. Virol. *78*, 1175-1184.

210. Mancini-Bourgine,M., Bayard,F., Soussan,P., Deng,Q., Lone,Y.C., Kremsdorf,D., and Michel,M.L. (2007). Hepatitis B virus splice-generated protein induces T-cell responses in HLA-transgenic mice and hepatitis B virus-infected patients. J. Virol. *81*, 4963-4972.

211. Marion,P.L., Oshiro,L.S., Regnery,D.C., Scullard,G.H., and Robinson,W.S. (1980). A virus in Beechey ground squirrels that is related to hepatitis B virus of humans. Proc. Natl. Acad. Sci. U. S. A *77*, 2941-2945.

212. McAleer,W.J., Buynak,E.B., Maigetter,R.Z., Wampler,D.E., Miller,W.J., and Hilleman,M.R. (1984). Human hepatitis B vaccine from recombinant yeast. Nature *307*, 178-180.

213. McCarthy, C. Chromas version 1.45. 1998.
Ref Type: Computer Program

214. McMahon,B.J., Alward,W.L., Hall,D.B., Heyward,W.L., Bender,T.R., Francis,D.P., and Maynard,J.E. (1985). Acute hepatitis B virus infection: relation of age to the clinical expression of disease and subsequent development of the carrier state. J Infect. Dis. *151*, 599-603.

215. Mehta,A., Lu,X., Block,T.M., Blumberg,B.S., and Dwek,R.A. (1997). Hepatitis B virus (HBV) envelope glycoproteins vary drastically in their sensitivity to glycan processing: evidence that alteration of a single N-linked glycosylation site can regulate HBV secretion. Proc. Natl. Acad. Sci U. S A *94*, 1822-1827.

216. 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 pre-S1 deleted. J. Virol. *71*, 5449-5454.

217. Meraldi,P., Lukas,J., Fry,A.M., Bartek,J., and Nigg,E.A. (1999). Centrosome duplication in mammalian somatic cells requires E2F and Cdk2-cyclin A. Nat. Cell Biol. *1*, 88-93.

218. Milich,D.R., Chen,M.K., Hughes,J.L., and Jones,J.E. (1998). The secreted hepatitis B precore antigen can modulate the immune response to the nucleocapsid: a mechanism for persistence. J. Immunol. *160*, 2013-2021.

219. 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. U. S. A *87*, 6599-6603.

220. Milich,D.R., McLachlan,A., Chisari,F.V., and Thornton,G.B. (1986). Nonoverlapping T and B cell determinants on an hepatitis B surface antigen pre-S(2) region synthetic peptide
4. J Exp. Med. *164*, 532-547.

221. Milich,D.R., McLachlan,A., Moriarty,A., and Thornton,G.B. (1987). A single 10-residue pre-S(1) peptide can prime T cell help for antibody production to multiple epitopes within the pre-S(1), pre-S(2), and S regions of HBsAg
4. J Immunol. *138*, 4457-4465.

222. Milich,D.R., McLachlan,A., Stahl,S., Wingfield,P., Thornton,G.B., Hughes,J.L., and Jones,J.E. (1988). Comparative immunogenicity of hepatitis B virus core and E antigens. J. Immunol. *141*, 3617-3624.

223. Mimms,L. (1995). Hepatitis B virus escape mutants: "pushing the envelope" of chronic hepatitis B virus infection. Hepatology *21*, 884-887.

224. Mizokami,M., Orito,E., Ohba,K., Ikeo,K., Lau,J.Y., and Gojobori,T. (1997). Constrained evolution with respect to gene overlap of hepatitis B virus. J. Mol. Evol. *44 Suppl 1*, S83-S90.

225. Moolla,N., Kew,M., and Arbuthnot,P. (2002). Regulatory elements of hepatitis B virus transcription. J. Viral Hepat. *9*, 323-331.

226. Moriyama,K., Takada,T., Tsutsumi,Y., Fukada,K., Ishibashi,H., Niho,Y., and Maeda,Y. (1994). Mutations in the transcriptional regulatory region of the precore and core/pregenome of a hepatitis B virus with defective HBeAg production. Fukuoka Igaku Zasshi *85*, 314-322.

227. Mulders,M.N., Venard,V., Njayou,M., Edorh,A.P., Bola Oyefolu,A.O., Kehinde,M.O., Muyembe Tamfum,J.J., Nebie,Y.K., Maiga,I., Ammerlaan,W., Fack,F., Omilabu,S.A., Ie,F.A., and Muller,C.P. (2004). Low genetic diversity despite hyperendemicity of hepatitis B virus genotype E throughout West Africa. J. Infect. Dis. *190*, 400-408.

228. Nagasaka,A., Hige,S., Marutani,M., Tsunematsu,I., Saito,M., Yamamoto,Y., Konishi,S., and Asaka,M. (1998). Prevalence of mutations in core promoter/precore region in Japanese patients with chronic hepatitis B virus infection. Dig. Dis. Sci. *43*, 2473-2478.

229. Nagasaki,F., Niitsuma,H., Cervantes,J.G., Chiba,M., Hong,S., Ojima,T., Ueno,Y., Bondoc,E., Kobayashi,K., Ishii,M., and Shimosegawa,T. (2006). Analysis of the entire nucleotide sequence of hepatitis B virus genotype B in the Philippines reveals a new subgenotype of genotype B. J. Gen. Virol. *87*, 1175-1180.

230. Nagaya,T., Nakamura,T., Tokino,T., Tsurimoto,T., Imai,M., Mayumi,T., Kamino,K., Yamamura,K., and Matsubara,K. (1987). The mode of hepatitis B virus DNA integration in chromosomes of human hepatocellular carcinoma. Genes Dev. *1*, 773-782.

231. Nakamoto,Y., Guidotti,L.G., Pasquetto,V., Schreiber,R.D., and Chisari,F.V. (1997). Differential target cell sensitivity to CTL-activated death pathways in hepatitis B virus transgenic mice 6. J Immunol. *158*, 5692-5697.

232. Nakao,K., Nakata,K., Yamashita,M., Tamada,Y., Hamasaki,K., Ishikawa,H., Kato,Y., Eguchi,K., and Ishii,N. (1999). p48 (ISGF-3gamma) is involved in interferon-alpha-induced suppression of hepatitis B virus enhancer-1 activity 3. J Biol. Chem. *274*, 28075-28078.

233. Nassal,M. and Schaller,H. (1993). Hepatitis B virus replication. Trends Microbiol. *1*, 221-228.

234. Nassal,M. and Schaller,H. (1996). Hepatitis B virus replication--an update. J. Viral Hepat. *3*, 217-226.

235. Naumann,H., Schaefer,S., Yoshida,C.F., Gaspar,A.M., Repp,R., and Gerlich,W.H. (1993). Identification of a new hepatitis B virus (HBV) genotype from Brazil that expresses HBV surface antigen subtype adw4. J. Gen. Virol. *74 (Pt 8)*, 1627-1632.

236. Neurath,A.R., Kent,S.B., Parker,K., Prince,A.M., Strick,N., Brotman,B., and Sproul,P. (1986a). Antibodies to a synthetic peptide from the preS 120-145 region of the hepatitis B virus envelope are virus neutralizing. Vaccine *4*, 35-37.

237. Neurath,A.R., Kent,S.B., and Strick,N. (1984). Location and chemical synthesis of a pre-S gene coded immunodominant epitope of hepatitis B virus. Science *224*, 392-395.

238. Neurath,A.R., Kent,S.B., Strick,N., and Parker,K. (1986b). Identification and chemical synthesis of a host cell receptor binding site on hepatitis B virus. Cell *46*, 429-436.

239. Neurath,A.R., Kent,S.B., Strick,N., Taylor,P., and Stevens,C.E. (1985). Hepatitis B virus contains pre-S gene-encoded domains. Nature *315*, 154-156.

240. Neurath,A.R., Seto,B., and Strick,N. (1989). Antibodies to synthetic peptides from the preS1 region of the hepatitis B virus (HBV) envelope (env) protein are virus-neutralizing and protective. Vaccine *7*, 234-236.

241. Ni,Y.H., Chang,M.H., Hsu,H.Y., and Chen,H.L. (2000). Long-term follow-up study of core gene deletion mutants in children with chronic hepatitis B virus infection. Hepatology *32*, 124-128.

242. Ni,Y.H., Huang,L.M., Chang,M.H., Yen,C.J., Lu,C.Y., You,S.L., Kao,J.H., Lin,Y.C., Chen,H.L., Hsu,H.Y., and Chen,D.S. (2007). Two decades of universal hepatitis B vaccination in taiwan: impact and implication for future strategies. Gastroenterology *132*, 1287-1293.

243. Nicholas, K. B and Nicholas, H. B. JR. GeneDoc: analysis and visualization of genetic variation. 1997.
Ref Type: Computer Program

244. Norder,H., Courouce,A.M., Coursaget,P., Echevarria,J.M., Lee,S.D., Mushahwar,I.K., Robertson,B.H., Locarnini,S., and Magnius,L.O. (2004). Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes. Intervirology *47*, 289-309.

245. 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.

246. Norder,H., Hammas,B., Lee,S.D., 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 (Pt 7)*, 1341-1348.

247. 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 (Pt 5)*, 1201-1208.

248. Norder,H., rauz-Ruiz,P., Blitz,L., Pujol,F.H., Echevarria,J.M., and Magnius,L.O. (2003). The T(1858) variant predisposing to the precore stop mutation correlates with one of two major genotype F hepatitis B virus clades. J. Gen. Virol. *84*, 2083-2087.

249. Nowak,M.A., Bonhoeffer,S., Hill,A.M., Boehme,R., Thomas,H.C., and McDade,H. (1996). Viral dynamics in hepatitis B virus infection 2. Proc. Natl. Acad. Sci U. S A *93*, 4398-4402.

250. Odemuyiwa,S.O., Mulders,M.N., Oyedele,O.I., Ola,S.O., Odaibo,G.N., Olaleye,D.O., and Muller,C.P. (2001). Phylogenetic analysis of new hepatitis B virus isolates from Nigeria supports endemicity of genotype E in West Africa. J. Med. Virol. *65*, 463-469.

251. Ohashi,R., Gao,C., Miyazaki,M., Hamazaki,K., Tsuji,T., Inoue,Y., Uemura,T., Hirai,R., Shimizu,N., and Namba,M. (2001). Enhanced expression of cyclin E and cyclin A in human hepatocellular carcinomas 5. Anticancer Res. *21*, 657-662.

252. Ohno,T., Mizokami,M., Wakita,T., and Kakumu,S. (1993). Defective short variants of S region in chronic HBV carriers. Nippon Rinsho *51*, 297-303.

253. Okamoto,H., Imai,M., Kametani,M., Nakamura,T., and Mayumi,M. (1987a). Genomic heterogeneity of hepatitis B virus in a 54-year-old woman who contracted the infection through materno-fetal transmission. Jpn. J Exp. Med. *57*, 231-236.

254. Okamoto,H., Imai,M., Tsuda,F., Tanaka,T., Miyakawa,Y., and Mayumi,M. (1987b). Point mutation in the S gene of hepatitis B virus for a d/y or w/r subtypic change in two blood donors carrying a surface antigen of compound subtype adyr or adwr. J Virol. *61*, 3030-3034.

255. Okamoto,H., Tsuda,F., Akahane,Y., Sugai,Y., Yoshiba,M., Moriyama,K., Tanaka,T., Miyakawa,Y., and Mayumi,M. (1994). Hepatitis B virus with mutations in the core promoter for an e antigen-negative phenotype in carriers with antibody to e antigen. J. Virol. *68*, 8102-8110.

256. Olinger,C.M., Venard,V., Njayou,M., Oyefolu,A.O., Maiga,I., Kemp,A.J., Omilabu,S.A., le,F.A., and Muller,C.P. (2006). Phylogenetic analysis of the precore/core gene of hepatitis B virus genotypes E and A in West Africa: new subtypes, mixed infections and recombinations. J. Gen. Virol. *87*, 1163-1173.

257. Ono-Nita,S.K., Kato,N., Shiratori,Y., Lan,K.H., Yoshida,H., Carrilho,F.J., and Omata,M. (1999a). Susceptibility of lamivudine-resistant hepatitis B virus to other reverse transcriptase inhibitors. J. Clin. Invest *103*, 1635-1640.

258. Ono-Nita,S.K., Kato,N., Shiratori,Y., Masaki,T., Lan,K.H., Carrilho,F.J., and Omata,M. (1999b). YMDD motif in hepatitis B virus DNA polymerase influences on replication and lamivudine resistance: A study by in vitro full-length viral DNA transfection. Hepatology *29*, 939-945.

259. 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. Japan HBV Genotype Research Group. Hepatology *33*, 218-223.

260. Orito,E., Sugauchi,F., Tanaka,Y., Ichida,T., Sata,M., Tanaka,E., Okanoue,T., Sakugawa,H., Watanabe,H., Miyakawa,H., Nishiguchi,S., Kumada,H., Ueda,R., and Mizokami,M. (2005). Differences of hepatocellular carcinoma patients with hepatitis B virus genotypes of Ba, Bj or C in Japan. Intervirology *48*, 239-245.

261. 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. U. S. A *83*, 1578-1582.

262. Ou,J.H. and Rutter,W.J. (1987). Regulation of secretion of the hepatitis B virus major surface antigen by the preS-1 protein. J Virol. *61*, 782-786.

263. Owiredu,W.K., Kramvis,A., and Kew,M.C. (2001a). 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.

264. Owiredu,W.K., Kramvis,A., and Kew,M.C. (2001b). Molecular analysis of hepatitis B virus genomes isolated from black African patients with fulminant hepatitis B. J. Med. Virol. *65*, 485-492.

265. Page, R. Treeview: An application to display phylogenetic trees on personal computers. 1996.
Ref Type: Computer Program

266. Papatheodoridis,G.V., Dimou,E., Laras,A., Papadimitropoulos,V., and Hadziyannis,S.J. (2002). Course of virologic breakthroughs under long-term lamivudine in HBeAg-negative precore mutant HBV liver disease. Hepatology *36*, 219-226.

267. Parekh,S., Zoulim,F., Ahn,S.H., Tsai,A., Li,J., Kawai,S., Khan,N., Trepo,C., Wands,J., and Tong,S. (2003). Genome replication, virion secretion, and e antigen expression of naturally occurring hepatitis B virus core promoter mutants. J. Virol. *77*, 6601-6612.

268. Parkin,D.M. (2001). Global cancer statistics in the year 2000. Lancet Oncol. *2*, 533-543.

269. Parkin,D.M., Pisani,P., and Ferlay,J. (1999). Estimates of the worldwide incidence of 25 major cancers in 1990. Int. J. Cancer *80*, 827-841.

270. Parkin,D.M., Sitas,F., Chirenje,M., Stein,L., Abratt,R., and Wabinga,H. (2008). Part I: Cancer in Indigenous Africans--burden, distribution, and trends. Lancet Oncol. *9*, 683-692.

271. 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.

272. Persson,B. and Argos,P. (1994). Prediction of transmembrane segments in proteins utilising multiple sequence alignments. J. Mol. Biol. *237*, 182-192.

273. Peterson,D.L., Nath,N., and Gavilanes,F. (1982). Structure of hepatitis B surface antigen. Correlation of subtype with amino acid sequence and location of the carbohydrate moiety
10. J Biol. Chem. *257*, 10414-10420.

274. 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.

275. Poussin,K., Dienes,H., Sirma,H., Urban,S., Beaugrand,M., Franco,D., Schirmacher,P., Brechot,C., and Paterlini,B.P. (1999). Expression of mutated hepatitis B virus X genes in human hepatocellular carcinomas. Int. J. Cancer *80*, 497-505.

276. Prange,R. and Streeck,R.E. (1995). Novel transmembrane topology of the hepatitis B virus envelope proteins. EMBO J. *14*, 247-256.

277. Preikschat,P., Gunther,S., Reinhold,S., Will,H., Budde,K., Neumayer,H.H., Kruger,D.H., and Meisel,H. (2002). Complex HBV populations with mutations in core promoter, C gene, and pre-S region are associated with development of cirrhosis in long-term renal transplant recipients. Hepatology *35*, 466-477.

278. Preikschat,P., Meisel,H., Will,H., and Gunther,S. (1999). Hepatitis B virus genomes from long-term immunosuppressed virus carriers are modified by specific mutations in several regions. J. Gen. Virol. *80 (Pt 10)*, 2685-2691.

279. Puisieux,A., Galvin,K., Troalen,F., Bressac,B., Marcais,C., Galun,E., Ponchel,F., Yakicier,C., Ji,J., and Ozturk,M. (1993). Retinoblastoma and p53 tumor suppressor genes in human hepatoma cell lines. FASEB J *7*, 1407-1413.

280. 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 1 binding site emerging in transplant-transmitted fulminant hepatitis B. Hepatology *25*, 1507-1515.

281. Qu,L., Huang,S., Baltzis,D., Rivas-Estilla,A.M., Pluquet,O., Hatzoglou,M., Koumenis,C., Taya,Y., Yoshimura,A., and Koromilas,A.E. (2004). Endoplasmic reticulum stress induces p53 cytoplasmic localization and prevents p53-dependent apoptosis by a pathway involving glycogen synthase kinase-3beta 4. Genes Dev. *18*, 261-277.

282. Radziwill,G., Tucker,W., and Schaller,H. (1990). Mutational analysis of the hepatitis B virus P gene product: domain structure and RNase H activity. J. Virol. *64*, 613-620.

283. 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.

284. Raney,A.K., Johnson,J.L., Palmer,C.N., and McLachlan,A. (1997). Members of the nuclear receptor superfamily regulate transcription from the hepatitis B virus nucleocapsid promoter. J. Virol. *71*, 1058-1071.

285. Rebbaa,A., Zheng,X., Chou,P.M., and Mirkin,B.L. (2003). Caspase inhibition switches doxorubicin-induced apoptosis to senescence. Oncogene *22*, 2805-2811.

286. Reddy,R.K., Mao,C., Baumeister,P., Austin,R.C., Kaufman,R.J., and Lee,A.S. (2003). Endoplasmic reticulum chaperone protein GRP78 protects cells from apoptosis induced by topoisomerase inhibitors: role of ATP binding site in suppression of caspase-7 activation 2. J Biol. Chem. *278*, 20915-20924.

287. Rehermann,B., Fowler,P., Sidney,J., Person,J., Redeker,A., Brown,M., Moss,B., Sette,A., and Chisari,F.V. (1995). The cytotoxic T lymphocyte response to multiple hepatitis B virus polymerase epitopes during and after acute viral hepatitis. J Exp. Med. *181*, 1047-1058.

288. Robinson,W.S. and Greenman,R.L. (1974). DNA polymerase in the core of the human hepatitis B virus candidate. J. Virol. *13*, 1231-1236.

289. Rosmorduc,O., Petit,M.A., Pol,S., Capel,F., Bortolotti,F., Berthelot,P., Brechot,C., and Kremsdorf,D. (1995). In vivo and in vitro expression of defective hepatitis B virus particles generated by spliced hepatitis B virus RNA. Hepatology *22*, 10-19.

290. Rozas,J. and Rozas,R. (1995). DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. Comput. Appl. Biosci. *11*, 621-625.

291. Sakamoto,T., Tanaka,Y., Orito,E., Co,J., Clavio,J., Sugauchi,F., Ito,K., Ozasa,A., Quino,A., Ueda,R., Sollano,J., and Mizokami,M. (2006). Novel subtypes (subgenotypes) of hepatitis B virus genotypes B and C

among chronic liver disease patients in the Philippines. J. Gen. Virol. *87*, 1873-1882.

292. 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.

293. Santantonio,T., Jung,M.C., Schneider,R., Fernholz,D., Milella,M., Monno,L., Pastore,G., Pape,G.R., and Will,H. (1992). Hepatitis B virus genomes that cannot synthesize pre-S2 proteins occur frequently and as dominant virus populations in chronic carriers in Italy 3. Virology *188*, 948-952.

294. Scaglioni,P., Melegari,M., Takahashi,M., Chowdhury,J.R., and Wands,J. (1996). Use of dominant negative mutants of the hepadnaviral core protein as antiviral agents. Hepatology *24*, 1010-1017.

295. Scaglioni,P.P., Melegari,M., and Wands,J.R. (1994). Characterization of hepatitis B virus core mutants that inhibit viral replication. Virology *205*, 112-120.

296. 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.

297. Schaefer,S. (2007a). Hepatitis B virus genotypes in Europe. Hepatol. Res. *37*, S20-S26.

298. Schaefer,S. (2007b). Hepatitis B virus taxonomy and hepatitis B virus genotypes. World J. Gastroenterol. *13*, 14-21.

299. Schirmacher,P., Wang,H., 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 Hepatol. *22*, 21-33.

300. Schlicht,H.J., Radziwill,G., and Schaller,H. (1989). Synthesis and encapsidation of duck hepatitis B virus reverse transcriptase do not require formation of core-polymerase fusion proteins. Cell *56*, 85-92.

301. Schluter,V., Meyer,M., Hofschneider,P.H., Koshy,R., and Caselmann,W.H. (1994). Integrated hepatitis B virus X and 3' truncated preS/S sequences derived from human hepatomas encode functionally active transactivators. Oncogene *9*, 3335-3344.

302. Seeger,C. and Mason,W.S. (2000). Hepatitis B virus biology. Microbiol. Mol. Biol. Rev. *64*, 51-68.

303. Seeger,C., Summers,J., and Mason,W.S. (1991). Viral DNA synthesis. Curr. Top. Microbiol. Immunol. *168*, 41-60.

304. Shafritz,D.A., Shouval,D., Sherman,H.I., Hadziyannis,S.J., and Kew,M.C. (1981). Integration of hepatitis B virus DNA into the genome of liver cells in chronic liver disease and hepatocellular carcinoma. Studies in percutaneous liver biopsies and post-mortem tissue specimens. N. Engl. J. Med. *305*, 1067-1073.

305. Shepard,C.W., Simard,E.P., Finelli,L., Fiore,A.E., and Bell,B.P. (2006). Hepatitis B virus infection: epidemiology and vaccination. Epidemiol. Rev. *28*, 112-125.

306. Shih,C., Burke,K., Chou,M.J., Zeldis,J.B., Yang,C.S., Lee,C.S., Isselbacher,K.J., Wands,J.R., and Goodman,H.M. (1987). Tight clustering of human hepatitis B virus integration sites in hepatomas near a triple-stranded region. J. Virol. *61*, 3491-3498.

307. Sirma,H., Giannini,C., Poussin,K., Paterlini,P., Kremsdorf,D., and Brechot,C. (1999). Hepatitis B virus X mutants, present in hepatocellular carcinoma tissue abrogate both the antiproliferative and transactivation effects of HBx. Oncogene *18*, 4848-4859.

308. Sommer,G., Gunther,S., Sterneck,M., Otto,S., and Will,H. (1997). A new class of defective hepatitis B virus genomes with an internal poly(dA) sequence. Virology *239*, 402-412.

309. Sommer,G., van,B.F., and Will,H. (2000). Genotype-specific synthesis and secretion of spliced hepatitis B virus genomes in hepatoma cells. Virology *271*, 371-381.

310. Soussan,P., Garreau,F., Zylberberg,H., Ferray,C., Brechot,C., and Kremsdorf,D. (2000). In vivo expression of a new hepatitis B virus protein encoded by a spliced RNA. J. Clin. Invest *105*, 55-60.

311. Standring,D.N., Ou,J.H., Masiarz,F.R., and Rutter,W.J. (1988). A signal peptide encoded within the precore region of hepatitis B virus directs the secretion of a heterogeneous population of e antigens in Xenopus oocytes. Proc. Natl. Acad. Sci. U. S. A *85*, 8405-8409.

312. Standring,D.N., Rutter,W.J., Varmus,H.E., and Ganem,D. (1984). Transcription of the hepatitis B surface antigen gene in cultured murine cells initiates within the presurface region. J. Virol. *50*, 563-571.

313. Steinhauer,D.A. and Holland,J.J. (1986). Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. J Virol. *57*, 219-228.

314. Stevens,C.E., Beasley,R.P., Tsui,J., and Lee,W.C. (1975). Vertical transmission of hepatitis B antigen in Taiwan. N. Engl. J Med. *292*, 771-774.

315. Stewart,L., Redinbo,M.R., Qiu,X., Hol,W.G., and Champoux,J.J. (1998). A model for the mechanism of human topoisomerase I. Science *279*, 1534-1541.

316. Stibbe,W. and Gerlich,W.H. (1983). Characterization of pre-s gene products in hepatitis B surface antigen
2. Dev. Biol. Stand. *54*, 33-43.

317. Stuyver,L., De,G.S., Van,G.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.

318. Sugauchi,F., Kumada,H., Acharya,S.A., Shrestha,S.M., Gamutan,M.T., Khan,M., Gish,R.G., Tanaka,Y., Kato,T., Orito,E., Ueda,R., Miyakawa,Y., and Mizokami,M. (2004a). Epidemiological and sequence differences between two subtypes (Ae and Aa) of hepatitis B virus genotype A. J. Gen. Virol. *85*, 811-820.

319. Sugauchi,F., Kumada,H., Sakugawa,H., Komatsu,M., Niitsuma,H., Watanabe,H., Akahane,Y., Tokita,H., Kato,T., Tanaka,Y., Orito,E., Ueda,R., Miyakawa,Y., and Mizokami,M. (2004b). Two subtypes of genotype B (Ba and Bj) of hepatitis B virus in Japan. Clin. Infect. Dis. *38*, 1222-1228.

320. Sugauchi,F., Mizokami,M., Orito,E., Ohno,T., Kato,H., Suzuki,S., Kimura,Y., Ueda,R., Butterworth,L.A., and Cooksley,W.G. (2001). A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness. J. Gen. Virol. *82*, 883-892.

321. Sugauchi,F., Ohno,T., Orito,E., Sakugawa,H., Ichida,T., Komatsu,M., Kuramitsu,T., Ueda,R., Miyakawa,Y., and Mizokami,M. (2003a). Influence of hepatitis B virus genotypes on the development of preS deletions and advanced liver disease. J. Med. Virol. *70*, 537-544.

322. Sugauchi,F., Orito,E., Ichida,T., Kato,H., Sakugawa,H., Kakumu,S., Ishida,T., Chutaputti,A., Lai,C.L., Ueda,R., Miyakawa,Y., and Mizokami,M. (2002). Hepatitis B virus of genotype B with or without recombination with genotype C over the precore region plus the core gene. J. Virol. *76*, 5985-5992.

323. Sugauchi,F., Orito,E., Kato,H., Suzuki,S., Kawakita,S., Sakamoto,Y., Fukushima,K., Akiba,T., Yoshihara,N., Ueda,R., and Mizokami,M. (2003b). Genotype, serotype, and phylogenetic characterization of the complete genome sequence of hepatitis B virus isolates from Malawian chronic carriers of the virus. J. Med. Virol. *69*, 33-40.

324. Summers,J. and Mason,W.S. (1982). Replication of the genome of a hepatitis B--like virus by reverse transcription of an RNA intermediate. Cell *29*, 403-415.

325. Summers,J., Smolec,J.M., and Snyder,R. (1978). A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchucks. Proc. Natl. Acad. Sci. U. S. A *75*, 4533-4537.

326. Sung,J.J., Tsui,S.K., Tse,C.H., Ng,E.Y., Leung,K.S., Lee,K.H., Mok,T.S., Bartholomeusz,A., Au,T.C., Tsoi,K.K., Locarnini,S., and Chan,H.L. (2008). Genotype-specific genomic markers associated with primary hepatomas, based on complete genomic sequencing of hepatitis B virus. J. Virol. *82*, 3604-3611.

327. Szmuness,W. (1978). Hepatocellular carcinoma and the hepatitis B virus: evidence for a causal association. Prog. Med. Virol. *24*, 40-69.

328. Szmuness,W., Prince,A.M., Grady,G.F., Mann,M.K., Levine,R.W., Friedman,E.A., Jacobs,M.J., Josephson,A., Ribot,S., Shapiro,F.L., Stenzel,K.H., Suki,W.N., and Vyas,G. (1974). Hepatitis B infection. A point-prevalence study in 15 US hemodialysis centers. JAMA *227*, 901-906.

329. Tai,P.C., Suk,F.M., Gerlich,W.H., Neurath,A.R., and Shih,C. (2002). Hypermodification and immune escape of an internally deleted middle-envelope (M) protein of frequent and predominant hepatitis B virus variants. Virology *292*, 44-58.

330. Takada,S. and Koike,K. (1990a). 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 *87*, 5628-5632.

331. Takada,S. and Koike,K. (1990b). X protein of hepatitis B virus resembles a serine protease inhibitor. Jpn. J. Cancer Res. *81*, 1191-1194.

332. Takada,S. and Koike,K. (1994). Three sites of the hepatitis B virus X protein cooperatively interact with cellular proteins. Virology *205*, 503-510.

333. Takahashi,K., Akahane,Y., Hino,K., Ohta,Y., and Mishiro,S. (1998). Hepatitis B virus genomic sequence in the circulation of hepatocellular carcinoma patients: comparative analysis of 40 full-length isolates. Arch. Virol. *143*, 2313-2326.

334. Takahashi,K., Aoyama,K., Ohno,N., Iwata,K., Akahane,Y., Baba,K., Yoshizawa,H., and Mishiro,S. (1995). The precore/core promoter mutant (T1762A1764) of hepatitis B virus: clinical significance and an easy method for detection. J. Gen. Virol. *76 (Pt 12)*, 3159-3164.

335. Takahashi,K., Ohta,Y., Kanai,K., Akahane,Y., Iwasa,Y., Hino,K., Ohno,N., Yoshizawa,H., and Mishiro,S. (1999). Clinical implications of mutations C-to-T1653 and T-to-C/A/G1753 of hepatitis B virus genotype C genome in chronic liver disease. Arch. Virol. *144*, 1299-1308.

336. Tamura, K, Dudley, J, Nei, M, and Kumar, S. MEGA5: Molecular Evolutionary Genetics Analysis. Center for Evolutionary Functional

GenomicsThe Biodesign Institute 1001 S.McAllister AvePO Box 875301Tempe, AZ 85287-5301 USAFax: 480-727-6947 . 2007. Ref Type: Electronic Citation

337. Tanaka,Y., Hasegawa,I., Kato,T., Orito,E., Hirashima,N., Acharya,S.K., Gish,R.G., Kramvis,A., Kew,M.C., Yoshihara,N., Shrestha,S.M., Khan,M., Miyakawa,Y., and Mizokami,M. (2004). A case-control study for differences among hepatitis B virus infections of genotypes A (subtypes Aa and Ae) and D. Hepatology *40*, 747-755.

338. Tanaka,Y., Mukaide,M., Orito,E., Yuen,M.F., Ito,K., Kurbanov,F., Sugauchi,F., Asahina,Y., Izumi,N., Kato,M., Lai,C.L., Ueda,R., and Mizokami,M. (2006). Specific mutations in enhancer II/core promoter of hepatitis B virus subgenotypes C1/C2 increase the risk of hepatocellular carcinoma. J. Hepatol. *45*, 646-653.

339. Tanaka,Y., Orito,E., Yuen,M.F., Mukaide,M., Sugauchi,F., Ito,K., Ozasa,A., Sakamoto,T., Kurbanov,F., Lai,C.L., and Mizokami,M. (2005). Two subtypes (subgenotypes) of hepatitis B virus genotype C: A novel subtyping assay based on restriction fragment length polymorphism. Hepatol. Res. *33*, 216-224.

340. Tang,J.J., Shen,C., and Lu,Y.J. (2006). Requirement for pre-existing of p21 to prevent doxorubicin-induced apoptosis through inhibition of caspase-3 activation
1. Mol. Cell Biochem. *291*, 139-144.

341. Terre,S., Petit,M.A., and Brechot,C. (1991). Defective hepatitis B virus particles are generated by packaging and reverse transcription of spliced viral RNAs in vivo. J. Virol. *65*, 5539-5543.

342. Testut,P., Renard,C.A., Terradillos,O., Vitvitski-Trepo,L., Tekaia,F., Degott,C., Blake,J., Boyer,B., and Buendia,M.A. (1996). A new hepadnavirus endemic in arctic ground squirrels in Alaska. J. Virol. *70*, 4210-4219.

343. Tiollais,P., Pourcel,C., and Dejean,A. (1985). The hepatitis B virus. Nature *317*, 489-495.

344. Tong,S. (2005). Mechanism of HBV genome variability and replication of HBV mutants. J. Clin. Virol. *34 Suppl 1*, S134-S138.

345. Torresi,J. (2002). The virological and clinical significance of mutations in the overlapping envelope and polymerase genes of hepatitis B virus. J. Clin. Virol. *25*, 97-106.

346. 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) with rearrangements in the pre-S/S and pre-C/C genes during chronic HBV infection. J. Virol. *65*, 3566-3574.

347. Tsebe,K.V., Burnett,R.J., Hlungwani,N.P., Sibara,M.M., Venter,P.A., and Mphahlele,M.J. (2001). The first five years of universal hepatitis B vaccination in South Africa: evidence for elimination of HBsAg carriage in under 5-year-olds. Vaccine *19*, 3919-3926.

348. Tu,H., Bonura,C., Giannini,C., Mouly,H., Soussan,P., Kew,M., Paterlini-Brechot,P., Brechot,C., and Kremsdorf,D. (2001). Biological impact of natural COOH-terminal deletions of hepatitis B virus X protein in hepatocellular carcinoma tissues. Cancer Res. *61*, 7803-7810.

349. Tuttleman,J.S., Pourcel,C., and Summers,J. (1986a). Formation of the pool of covalently closed circular viral DNA in hepadnavirus-infected cells
1. Cell *47*, 451-460.

350. Tuttleman,J.S., Pugh,J.C., and Summers,J.W. (1986b). In vitro experimental infection of primary duck hepatocyte cultures with duck hepatitis B virus
2. J Virol. *58*, 17-25.

351. Vogelstein,B., Lane,D., and Levine,A.J. (2000a). Surfing the p53 network
1. Nature *408*, 307-310.

352. Vogelstein,B., Lane,D., and Levine,A.J. (2000b). Surfing the p53 network
1. Nature *408*, 307-310.

353. von Weizsacker F., Kock,J., Wieland,S., Offensperger,W.B., and Blum,H.E. (1999). Dominant negative mutants of the duck hepatitis B virus core protein interfere with RNA pregenome packaging and viral DNA synthesis. Hepatology *30*, 308-315.

354. von Weizsacker F., Wieland,S., and Blum,H.E. (1996). Inhibition of viral replication by genetically engineered mutants of the duck hepatitis B virus core protein. Hepatology *24*, 294-299.

355. Wagayama,H., Shiraki,K., Yamanaka,T., Sugimoto,K., Ito,T., Fujikawa,K., Takase,K., and Nakano,T. (2001). p21WAF1/CTP1 expression and hepatitis virus type
1. Dig. Dis. Sci *46*, 2074-2079.

356. Wakita,T., Kakumu,S., Shibata,M., Yoshioka,K., Ito,Y., Shinagawa,T., Ishikawa,T., Takayanagi,M., and Morishima,T. (1991). Detection of pre-C and core region mutants of hepatitis B virus in chronic hepatitis B virus carriers. J. Clin. Invest *88*, 1793-1801.

357. Wang,G.H. and Seeger,C. (1992). The reverse transcriptase of hepatitis B virus acts as a protein primer for viral DNA synthesis. Cell *71*, 663-670.

358. Wang,H.C., Chang,W.T., Chang,W.W., Wu,H.C., Huang,W., Lei,H.Y., Lai,M.D., Fausto,N., and Su,I.J. (2005). Hepatitis B virus pre-S2 mutant upregulates cyclin A expression and induces nodular proliferation of hepatocytes. Hepatology *41*, 761-770.

359. Wang,H.C., Huang,W., Lai,M.D., and Su,I.J. (2006). Hepatitis B virus pre-S mutants, endoplasmic reticulum stress and hepatocarcinogenesis
12. Cancer Sci *97*, 683-688.

360. Wang,H.C., Wu,H.C., Chen,C.F., Fausto,N., Lei,H.Y., and Su,I.J. (2003). Different types of ground glass hepatocytes in chronic hepatitis B virus infection contain specific pre-S mutants that may induce endoplasmic reticulum stress. Am. J. Pathol. *163*, 2441-2449.

361. Wang,H.P. and Rogler,C.E. (1991). Topoisomerase I-mediated integration of hepadnavirus DNA in vitro. J Virol. *65*, 2381-2392.

362. Wang,J.C. (1985). DNA topoisomerases. Annu. Rev. Biochem. *54*, 665-697.

363. Wang,W.X., Li,M., Wu,X., Wang,Y., and Li,Z.P. (1998). HNF1 is critical for the liver-specific function of HBV enhancer II. Res. Virol. *149*, 99-108.

364. Wang,Y., Chen,P., Wu,X., Sun,A.L., Wang,H., Zhu,Y.A., and Li,Z.P. (1990). A new enhancer element, ENII, identified in the X gene of hepatitis B virus. J. Virol. *64*, 3977-3981.

365. Wang,Y., Lau,S.H., Sham,J.S., Wu,M.C., Wang,T., and Guan,X.Y. (2004). Characterization of HBV integrants in 14 hepatocellular carcinomas: association of truncated X gene and hepatocellular carcinogenesis. Oncogene *23*, 142-148.

366. 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.

367. WHO. http://www.who.int. Hepatitis B.World Health Organization Department of Communicable diseases Surveillance Reponse.WHO/CDS/CSR/LYO/2002.2: Hepatits B. 2002.
Ref Type: Electronic Citation

368. WHO. Global Immunization Data. http://www.who.int/immunization/newsroom/Global_Immunization_Data.pdf. World Health Organization . 2008.
Ref Type: Electronic Citation

369. 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.

370. Wu,H.L., Chen,P.J., Tu,S.J., Lin,M.H., Lai,M.Y., and Chen,D.S. (1991). Characterization and genetic analysis of alternatively spliced transcripts of hepatitis B virus in infected human liver tissues and transfected HepG2 cells. J Virol. *65*, 1680-1686.

371. Xiong,Y., Hannon,G.J., Zhang,H., Casso,D., Kobayashi,R., and Beach,D. (1993). p21 is a universal inhibitor of cyclin kinases. Nature *366*, 701-704.

372. Xu,Z. and Yen,T.S. (1996). Intracellular retention of surface protein by a hepatitis B virus mutant that releases virion particles. J Virol. *70*, 133-140.

373. Yaginuma,K., Kobayashi,H., Kobayashi,M., Morishima,T., Matsuyama,K., and Koike,K. (1987). Multiple integration site of hepatitis B virus DNA in hepatocellular carcinoma and chronic active hepatitis tissues from children. J. Virol. *61*, 1808-1813.

374. Yamamoto,K., Horikita,M., Tsuda,F., Itoh,K., Akahane,Y., Yotsumoto,S., Okamoto,H., Miyakawa,Y., and Mayumi,M. (1994). Naturally occurring escape mutants of hepatitis B virus with various mutations in the S gene in carriers seropositive for antibody to hepatitis B surface antigen. J. Virol. *68*, 2671-2676.

375. Yen,T.S. (1996). Hepadnaviral X Protein:Review of Recent Progress. J Biomed. Sci *3*, 20-30.

376. Yu,M.W., Yeh,S.H., Chen,P.J., Liaw,Y.F., Lin,C.L., Liu,C.J., Shih,W.L., Kao,J.H., Chen,D.S., and Chen,C.J. (2005). Hepatitis B virus genotype and DNA level and hepatocellular carcinoma: a prospective study in men. J. Natl. Cancer Inst. *97*, 265-272.

377. Yu,X. and Mertz,J.E. (1996). Promoters for synthesis of the pre-C and pregenomic mRNAs of human hepatitis B virus are genetically distinct and differentially regulated. J. Virol. *70*, 8719-8726.

378. Yuen,M.F., Tanaka,Y., Mizokami,M., Yuen,J.C., Wong,D.K., Yuan,H.J., Sum,S.M., Chan,A.O., Wong,B.C., and Lai,C.L. (2004). Role of hepatitis B virus genotypes Ba and C, core promoter and precore mutations on hepatocellular carcinoma: a case control study. Carcinogenesis *25*, 1593-1598.

379. 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.

380. Yuh,C.H. and Ting,L.P. (1991). C/EBP-like proteins binding to the functional box-alpha and box-beta of the second enhancer of hepatitis B virus. Mol. Cell Biol. *11*, 5044-5052.

381. Yuh,C.H. and Ting,L.P. (1993). Differentiated liver cell specificity of the second enhancer of hepatitis B virus. J. Virol. *67*, 142-149.

382. Zeuzem,S., de Man,R.A., Honkoop,P., Roth,W.K., Schalm,S.W., and Schmidt,J.M. (1997). Dynamics of hepatitis B virus infection in vivo 2. J Hepatol. *27*, 431-436.

383. Zhang,P. and McLachlan,A. (1994). Differentiation-specific transcriptional regulation of the hepatitis B virus nucleocapsid gene in human hepatoma cell lines. Virology *202*, 430-440.

384. Zhang,P., Raney,A.K., and McLachlan,A. (1993). Characterization of functional Sp1 transcription factor binding sites in the hepatitis B virus nucleocapsid promoter. J. Virol. *67*, 1472-1481.

385. Ziemer,M., Garcia,P., Shaul,Y., and Rutter,W.J. (1985). Sequence of hepatitis B virus DNA incorporated into the genome of a human hepatoma cell line. J Virol. *53*, 885-892.