CHAPTER 1

1.0. INTRODUCTION

Hepatitis B virus (HBV) infection is endemic in several parts of the world. It is responsible for a series of diseases ranging from acute hepatitis, uncomplicated or fulminant, through chronic hepatitis to cirrhosis and hepatocellular carcinorma. The virus may also exist in a chronic asymptomatic carrier state. It accounts for about 1 million deaths each year, predominantly as a result of disease progression to cirrhosis, hepatocellular carcinorma, (HCC) or both diseases (Alberti *et al*., 2005). Approximately 385 million people worldwide are asymptomatically infected with the virus and these people are at a high risk of developing HCC (Kane, 1995). Africa is one of the most affected continents with about 50 million chronically infected individuals (figure 1.1) (Kiire, 1996, Seeger and Mason, 2000). Africa also accounts for about one quarter of a million deaths each year (Kew, 1992).

The carrier rate in Africa ranges from 6.5% in Tunisia, north Africa to 19% in Niger, west Africa, with an average of 10.4% throughout the continent (Kew, 1992). In sub-Saharan Africa the virus is hyperendemic, with a carrier rate of 5-20 % (Mphahlele *et al*., 2002)**.** The carrier rate among men is slightly higher than in women, averaging 1.4-2.5:1. The lifelong HBV exposure rates (individuals exposed to HBV at any time during life) do not differ between sexes and ranges from 57% in Nigeria, west Africa to 98% in Namibia, southern Africa (Kew, 1992).

The comparison of urban and rural data in terms of carrier rates and lifelong exposure is of vital importance to the epidemiology of HBV infection, especially in Africa, in which the population is mostly rural. South Africa is the greatest contributor of such data in Africa. In one of such studies carried out in the Transkei (coastal region of South Africa) the chronic HBV infection rate in rural areas was double that in urban areas (15.5% and 7.4%, respectively) (Vos *et al*., 1984).

Initial HBV infection in endemic regions of the world (Africa and the Far East) occurs mostly during birth (perinatal transmission) or early childhood (horizontal transmission). Horizontal transmission is the predominant route of infection in Africa, whereas perinatal infection in the predominant route in the Far East. Perinatal transmission occurs mainly in children born to mothers with replicative HBV infection. The number of pregnant women with replicative HBV infection is lower in Africa than the Far East (14% and 40%, respectively) (Kew, 1992). The difference in the persistence of the replicative virus in Asian mothers may be the result of infection with different genotypes from those infecting African mothers (Kramvis *et al*., 2005).

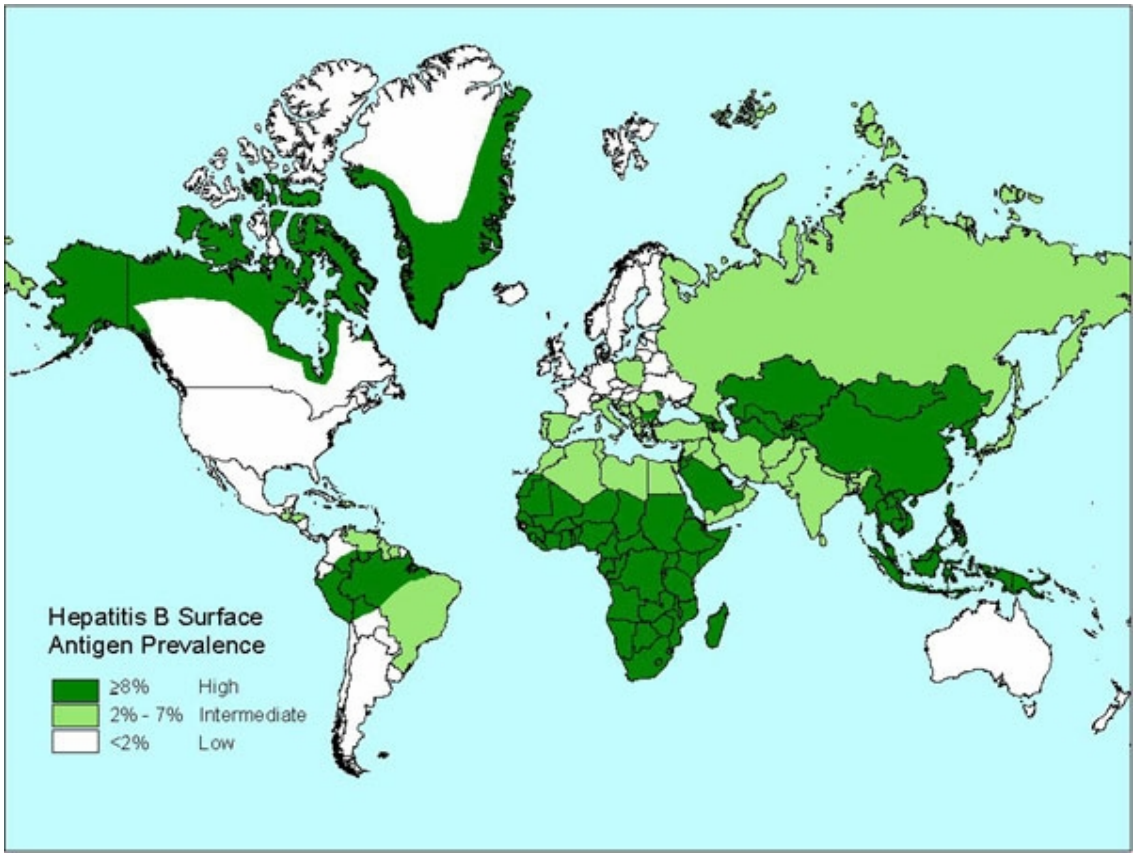


Figure 1.1. Geographical distribution of chronic hepatitis B virus infection. http://www.who.int**.**

There are limited data on complete sequences of HBV genomes from southern Africa. Complete genome analysis is important in the study of HBV variants because HBV usually circulates as a quasispecies (more than one variant at a time) (Blum, 1993). Hence, data based on sub-genomic amplification does not indicate whether one or more viral genomes are being characterized. To characterize and functionally determine the effect of mutations it is important to study single complete viral genomes, as sub-genomic amplification may fail to identify mutations on the same DNA molecule.

The differences in the geographic distribution of the genotypes of HBV and of the population groups in the world make it difficult to extrapolate data obtained from other regions of the world to the southern African context. Therefore, it is important that the strains circulating in the southern African population are studied in depth, in order to provide an insight into HBV infection and its clinical manifestations in this geographic region.

HBV can be divided into 8 genotypes, which show a distinct geographical distribution. Analysis of a small number of sub-genomic and complete genome sequences from South Africa has lead to the identification of two predominant HBV genotypes from this region (genotype A and D) (Bowyer *et al*., 1997, Owiredu *et al*., 2001a, and Owiredu *et al*., 2001b). These genotypes have unique mutations in the various open reading frames as well as *cis*-acting elements, which may account for differences in disease presentations in this region.

The main objective of this study was to augment the database of complete HBV genome sequences from South Africa and to functionally characterize some of the unique mutations present in HBV genotypes from this region.

1.1. THE FAMILY *HEPADNAVIRIDAE*

The family *Hepadnaviridae* is divided into 2 genera; the Orthohepadnaviruses and the Avihepadnaviruses. HBV is the prototype of the Orthohepadnaviruses. These include the woodchuck hepatitis virus (WHV) infecting eastern woodchucks (*Marmota Monax*), ground squirrel hepatitis virus (GSHV) infecting the California beechy ground squirrels (*Spermophilus beecheyi*), Arctic squirrel hepatitis virus (ASHV) infecting the wild artic ground squirrels (*Spermophilus parryi kennicotti*), and woolly monkey hepatitis virus (WMHBV) infecting south American woolly monkeys (*Lagothrix lagotricha*). WMHBV is the most closely related of the Orthohepadnaviruses to HBV (Summers *et al*., 1978, Marion *et al*., 1980, Testut *et al*., 1996, Lanford *et al*., 1998). Non-human primates, such as chimpanzees (*Pan pygmeaeus*), white-handed gibbons (*Hybolates lar*), orangutans (*Pongo pygmeaeus*), and gorillas (*Gorilla gorilla*), may also be infected with hepadnaviruses (Vaudin *et al*., 1988, Norder *et al*., 1996, Warren *et al*., 1999, Grethe *et al*., 2000, Lanford *et al*., 2000, Verschoor *et al*., 2001).

The Avihepadnaviruses consist of duck HBV (DHBV), which infects Pekin ducks (*Anas platyrhynschus*) and geese (*Anser indicus*), the heron HBV (HHBV) that is found in German grey herons (*Ardea cinerea*), and white stork HBV (STHBV) that infects whites storks (*Ciconia ciconia*) (Marion *et al*., 1980, Mason *et al*., 1980, Pult *et al*., 2001, Sprengel *et al*., 1988, Summers *et al*., 1978). The members of the family *Hepadnaviridae* are highly hepatotropic and cause persistent infection (Robinson *et al*., 1987, Zlotnick *et al*., 1998). One important feature of the hepadnaviruses is the ability to produce infectious and/or non-infectious forms of the virus in the blood during the

course of infection (Seeger and Mason, 2000). Some members of the Orthohepadnaviruses may give rise to HCC.

1.1.1. VIRAL STRUCTURE

HBV exists in serum in complete (Dane particle) and incomplete viral forms. The complete viral particle (virion) is spherical with a diameter of 42nm (figure 1.2a). It is composed of an outer envelope and an inner icosahedral nucleocapsid. The envelope is made up of large (LHBs), middle (MHBs), and small (SHBs) surface proteins. The nucleocapsid is composed of the core antigen (HBcAg), which harbors the viral DNA and DNA dependent polymerase (Almeida, 1971, Bayer *et al*., 1968, Dane *et al*., 1970, Robinson *et al*., 1974, Robinson and Greenman, 1974, Seeger and Mason, 2000, Seeger *et al*., 1991). The incomplete viral forms are either spherical or filamentous. They contain no nucleic acid, and are made up of proteins, carbohydrates, and lipids. They bear the HBV surface antigen (HBsAg) on their surface. The spherical forms are about 22nm in diameter, whereas filamentous forms are 22nm wide with varying lengths (figure 1.2b and figure 1.2c)

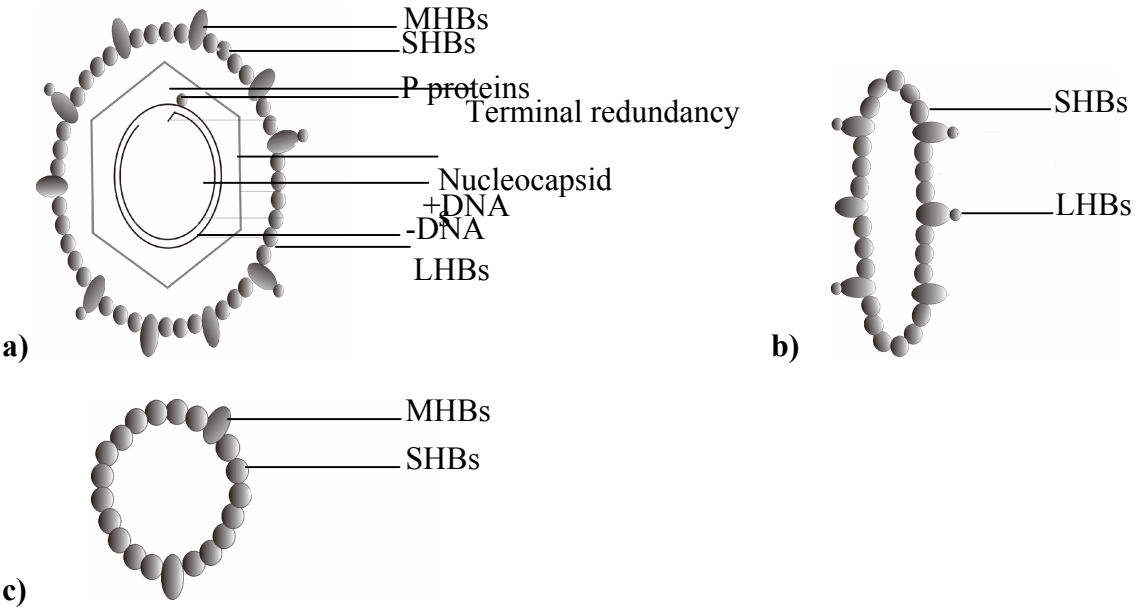


Figure 1.2. a) Diagrammatic representation of the complete HBV virion (Dane particle). The envelope is composed of the LHBs, MHBs, and SHBs. The LHBs is made up of the pre-S1, pre-S2, and S domains, the MHBs pre-S2 and S, and the SHBs the S. The DNA is located within the icosahedral nucleocapsid. P is the DNA polymerase covalently linked to the 5' end of the minus strand b) Incomplete filamentous viral form c) Incomplete spherical form.

1.1.2. VIRAL GENOME

HBV has the smallest genome of all known animal viruses (Feitelson, 1994). The genome of the virus is made up of a partially double-stranded DNA molecule (Robinson *et al*., 1974) (figure 1.2a). It consists of a complete minus strand that is about 3200 base pairs (bp) in length and an incomplete positive strand, which may contain between 1700 and 2800 bp (Hruska *et al*., 1977, Landers *et al*., 1977, Summers *et al*., 1975). The minus strand encodes for the viral proteins and pregenome, whereas the

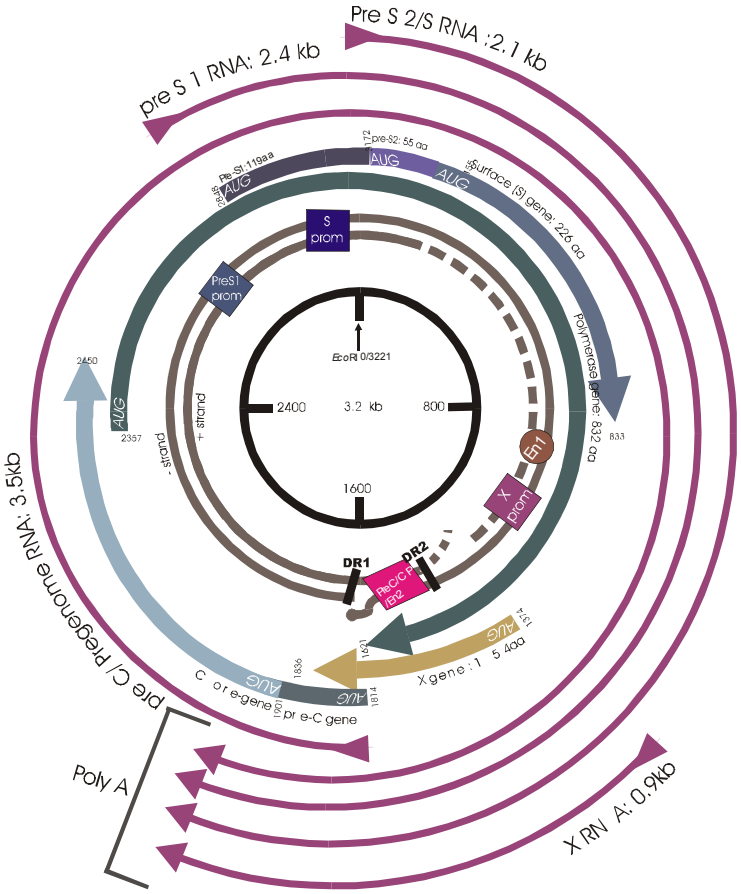
positive strand is non-coding (Galibert *et al*., 1979, Miller *et al*., 1989). The positive strand is prevented from phosphorylation by a capped oligoribonucleotide, covalently attached to its 5' end, and the negative strand is blocked by a polypeptide (DNA polymerase) covalently linked to it 5'end (Gerlich and Robinson, 1980, Seeger *et al*., 1986). The circular nature of the viral DNA is maintained by base pairing at the 5' ends of both strands (the cohesive overlap region) (Tiollais *et al*., 1985).

1.1.3. HBV TRANSCRIPTS

HBV generates 4 categories of unidirectional and overlapping transcripts from 4 open reading frames (ORFs) and these are translated into 7 proteins products (figure 1.3). These transcripts are produced from covalently closed circular DNA (cccDNA) viral genomes with the aid of the host transcription machinery. cccDNA synthesis takes place in the nucleus during the viral life cycle. This involves the removal of nucleotides from the 5' end of the plus strand and redundant nucleotides from the minus strand, followed by ligation of the ends of both strands by the endogenous host polymerase (Summers and Mason, 1982). The start sites are distributed across the circular DNA viral genome with a common polyadenylation signal (table 1.1).

The longest and the most abundant transcript, the 3.5 kb transcript, can be separated into two subsets of functionally distinct mRNAs (precore mRNA and the core mRNA or pregenome) (Yaginuma *et al*., 1989). The precore mRNA serves as a template for the synthesis of the HBV e antigen (HBeAg), while the core mRNA serves as the template for reverse transcription of the genome, as well as core protein and polymerase

synthesis. The transcription start of the precore mRNA is heterogeneous (either starting at 1783 or 1784) (Yaginuma *et al*., 1989), whereas that of the core mRNA is homogeneous, beginning at 1818 (Will *et al*., 1987). The LHBs protein is translated from the 2.4 kb mRNA known as the pre-S1 transcript, whereas the 2.1 kb mRNA (pre-S2 transcript) is responsible for the expression of the MHBs and HBsAg. The smallest transcript is a 0.7 kb mRNA, which is responsible for HBx protein production (Nassal and Schaller, 1993).



preC/C P/En1 =precore/core promoter and enhancer I, X prom= X promoter En2= enhancer II, S prom= S promoter, Pre-S1 prom= Pre-S1 promoter

Figure 1.3. Organization of the HBV genome. Numbering is from the *Eco*RI cleavage site. The partially dashed inner circle represents the incomplete plus-strand where the reverse transcriptase attaches to the 5' end. The minus strand is also capped at its 5' end by a RNA oligomer. The positions of the viral *cis*-elements are represented by rectangular and circular symbols. The immediate surrounding arrows represent the viral open reading frames. The four outer arrows indicate the HBV transcripts. All transcripts share a common poly-adenylation site.

Table 1.1. HBV transcripts

Size (kb)	Start sites[a]	Calculated length[b] (kb)	Function
3.5	1783/1784 or 1790±1	3.5	PreCore mRNAs
3.5	1818	3.3	Core mRNA/ Pregenome
2.4	2807	2.3	Pre-S1 mRNA
2.1	3157/3158	2.0	Pre-S2 mRNA
2.1	3174/3175	1.95	SHBs mRNA
	5/6		
	113		
0.7	1226/1242	0.65-0.7	XmRNA

[a] Nucleotide numbering according to Galibert *et al*., 1979.

[b]Approximate length without poly A tail.

Data compiled from:Yaginuma *et al*., 1987, Will *et al*., 1987., Cattaneo *et al*., 1984, and Standring *et al*., 1984.

1.1.4 VIRAL GENE PRODUCTS

The HBV genome is made up of 4 overlapping open reading frames (ORFs) encoding 7 viral proteins (Tiollais *et al*., 1985) (figure 1.3). The compact organization of the HBV genome limits the ability of the virus to tolerate mutations, and when mutations do occur they affect the virus significantly (Mizokami *et al*., 1997).

1.1.4.1 PRECORE AND CORE GENE PRODUCTS

The core and precore mRNA encode for the HBcAg and HBeAg, respectively, (Tiollais *et al*., 1985). A proportion of the core protein is translated from the precore mRNA. This, together with that translated from core mRNAs, make up the nucleocapsid (Chang

et al., 1994, Tiollais *et al*., 1985). Translation from the outer precore initiation site results in the formation of a 210 amino acids (aa) long, HBeAg precursor (p25). The N-terminal aa residues of p25 (-29 to –11) target the nascent protein to the endoplasmic reticulum, where it is cleaved to generate p23/p22. The C-terminal arginine rich region (34 aa long) is subsequently removed by a furin-like protease in the Golgi apparatus to yield p17 (the mature HBeAg), which is then secreted into the blood stream or expressed on the surface of the hepatocytes (Bruss and Gerlich, 1988, Ou *et al*., 1986, Standring *et al*., 1988). HBeAg is thought to play a role in the persistence of HBV infection (Milich *et al*., 1998, Milich *et al*., 1990) by regulating the immune response to the core antigen (Chen *et al*., 2004).

1.1.4.2. SURFACE PROTEINS

The infectious virion contains an envelope that is made up of three types of surface proteins embedded in a lipid membrane: LHBs, MHBs, and SHBs/HBsAg (figure1.2a). The pre-S1 mRNA transcript codes for the LHBs, whereas the pre-S2 codes for MHBs and SHBs/HBsAg. The 3 surface proteins have a common stop codon and three different in-frame start sites (Tiollais *et al*., 1985). Seventy percent of the complete virion envelope is made up of SHBs/HBsAg. MHBs and LHBs comprise the remaining portion in equal proportions. The non-infectious spherical particles are composed mainly of SHBs/HBsAg, whereas the filamentous particle contain mainly SHBs/HBsAg but also MHBs and LHBs in small amounts (Heermann *et al*., 1984, Heermann *et al*., 1987, Seeger and Mason, 2000).

The general function of the surface proteins seems to be the protection of the nucleocapsid from lysis during transportation of the virus from the hepatocyte. The pre-S1 domain has two main functions: binding of the Dane particle to the hepatocyte during infection and providing a ligand for the core particle during assembly of the viral envelope (Seeger and Mason, 2000). The pre-S2 domain is dispensable for HBV infection of hepatocytes (Fernholz *et al.*, 1993), but has transctivating activity when truncated on the carboxyl terminal (Hildt *et al*., 1996)

HBsAg is made up of at least three hydrophobic α-helices with two hydrophilic regions connecting them. The first α-helix spanning aa 11-29, is located within the endoplasmic reticulum (ER) membrane and functions in the translocation of upstream sequences (Eble *et al*., 1990) and the second α-helix, spanning residues 80-98, inhibits translocation of upstream sequences while initiating translocation of downstream sequences (Eble *et al*., 1987). The function of the third α-helix spanning residues168-186 is unknown. Amino acids 1-10 and the second hydrophilic region aa 99-167 are located in-face with the lumen of the endoplasmic reticulum, while the first hydrophilic region aa 30-79 faces the cytoplasm. The carboxyl-terminal aa 210-226 is exposed on the cell surface (Gerlich *et al*., 1992).

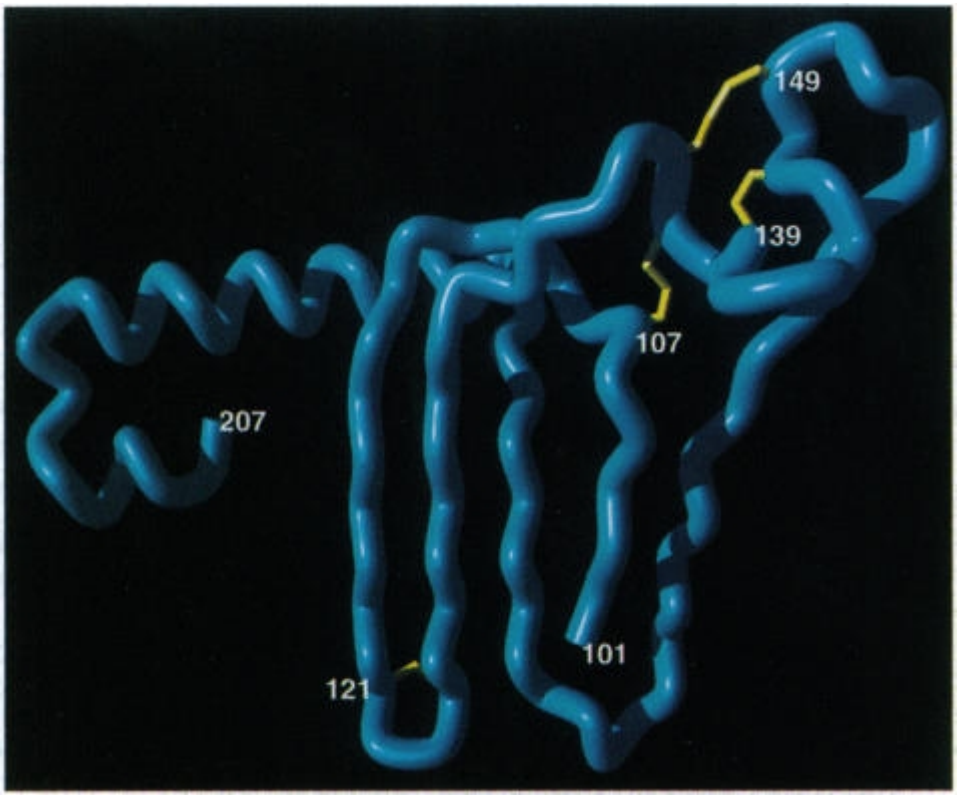


Figure 1.4. Model of the HBsAg showing the second hydrophilic region. The antigenic determinant 'a' is found between aa 138 and 149 (top right hand corner). Disulphide bonds are marked in yellow and are formed between aa 120 and 124, 107 and 138, 137 and 149, and 139 and 147 (Chen *et al*., 1996).

The region between the second and the third hydrophilic regions represents classical HBsAg and contains the antigenic 'a' determinant, common to all serological subtypes. The 'a' determinant is a cyclic structure of high antigenicity located between aa 138-149. It is stabilized by disulphide bonds between the six cysteine residues (Gerlich *et al*., 1992).

1.1.4.3. POLYMERASE

The viral DNA polymerase is a multifunctional 95kDa protein encoded by the polymerase ORF (P-ORF). It spans two-thirds of the viral genome, overlapping with the other three ORFs (Gerlich *et al*., 1992, Radziwill *et al*., 1990). The polymerase is transcribed from a start point internal to the pregenome or core mRNA (Chang *et al*., 1990, Chang *et al*., 1989, Schlicht *et al*., 1989), and it therefore lacks it own promoter elements, leaving its regulation mainly at the level of translation (Hwang and Su, 1998).

Viral DNA polymerase is made up of four domains (Radziwill *et al*., 1990). It lacks the intergrase-like domain present in retroviruses, indicating that integration is not necessary for replication (Radziwill *et al*., 1990). The first functional domain, the terminal protein (TP) domain, is found at the amino terminal of the polymerase protein. The TP spans the region between aa 1-183. The TP domain is covalently linked to the minus strand DNA at aa 63, preventing phosphorylation of the minus strand (Bartenschlager and Schaller, 1988, Bosch *et al*., 1988). Its other functions include priming and synthesis of the minus strand and packaging of pregenomic RNA (Seeger and Mason, 2000, Seeger *et al*., 1991, Wang and Seeger, 1992). The next domain, the spacer domain comprising aa 184-354, separates the TP domain from the reverse transcriptase domain. It has been shown to be highly tolerant to mutations and hence varies greatly between different HBV genotypes (Radziwill *et al*., 1990). The third domain, the reverse transcriptase (RT) domain comprising aa 355-694, contains the YMDD consensus motif. It functions in reverse transcription and priming. The last

domain, the RNaseH domain comprising aa 695- 845, is present at the carboxyl-terminal end of the polymerase protein, and most probably functions in stabilizing the TP-RT complex during interaction with the template and in the degradation of the 17-18 bp residual pregenome during elongation of the minus strand (Radziwill *et al*., 1990).

1.1.4.4. HBx PROTEIN (X PROTEIN)

The X protein, a 17 kD protein, is encoded by the 475 nucleotide long X ORF (Kay *et al*., 1985, Tiollais *et al*., 1985). The X protein has been localized both in the cytoplasm and the nucleus of infected hepatocytes (Henkler *et al*., 2001, Hoare *et al*., 2001). HBx is necessary for establishment of HBV infection (Chen *et al*., 1993) and has been implicated in the development of HCC (Koike *et al*., 1989).

HBx possesses transactivating activity. This allows it to regulate the transcription of several cellular and viral genes. Because the X protein lacks DNA binding ability (Seeger and Mason, 2000), its transactivating activity is most probably carried out by interaction with cellular proteins. This interaction is facilitated by the presence of a Kunitz domain-like site on the HBx. Mutations within this site have been shown to abolish HBx transactivating activity. The other functional domains of the HBx are the amino terminal domain, which down-regulates its transactivating activity, and the histidine/asparagine rich domain, which may play a role in transactivation (Arii *et al*., 1992, Takada and Koike, 1990).

1.1.5. VIRAL *CIS*- ELEMENTS

1.1.5.1. PROMOTERS AND ENHANCERS

Four independent promoters and two enhancers control transcription of six transcripts of HBV. The core promoter encompasses a 232 bp region, which is located upstream from the core gene, and overlaps with the X and precore gene. It directs the transcription of the 3.5 kb transcripts, precore mRNA and core mRNA or pregenomic RNA (pgRNA). Its functional units include the basic core promoter (BCP) and an upstream regulatory region (URR). The BCP lacks the TATA box characteristic of most promoter sequences. The URR is made up of a core upstream regulatory sequence (CURS), which is involved in activation of the BCP and a negative regulatory element, which suppresses the activity of BCP, and enhancer II (EnhII) (Guo *et al*., 1991, Kramvis and Kew, 1999, Lo and Ting, 1994, Yu and Mertz, 1996, Yuh *et al*., 1992).

The pre-S1 promoter controls the transcription of the pre-S1 mRNA (2.4 kb transcript) that is translated into LHBs. On the other hand, the transcription of the pre-S2 mRNA (2.1 kb), which is translated into MHBs and HBsAg, is directed by the pre-S2 promoter. The fourth promoter, the X promoter, directs the transcription of the X mRNA. (Cattaneo *et al*., 1983, Schaller and Fischer, 1991).

The two HBV enhancers function by activating different HBV promoters. The enhancer I (EnhI) is located between nucleotide 1074-1234 and functions in activating the pre-S1, pre-S2, core and X promoters in a differentiated-hepatocyte specific manner (Antonucci and Rutter, 1989, Guo *et al*., 1991, Hu and Siddiqui, 1991, Schaller and

Fischer, 1991). The enhancer II maps at nucleotide 1636-1741. It functions in up-regulating the X and surface promoters and activating the BCP. It is made up of two functional elements, A and B. The interaction of these two elements is essential for its activity, with functional element B accounting for about three quarters of its function (Yuh and Ting, 1990, Yuh and Ting, 1993). A glucocorticoid-responsive element (GRE) is located within the S ORF, a few nucleotides upstream from nucleotide 735 (Tur-Kaspa *et al*., 1986). The GRE is distinct from the enhancer elements and mutations within it can affect response to glucocorticoids (Suzuki *et al*., 1998).

1.1.5.2. THE ENCAPSIDATION SIGNAL

The encapsidation signal, epsilon (ε), is a 70-nucleotide stem-loop structure present at the 5' end of the pgRNA (Junker-Niepmann *et al*., 1990). It is required for viral encapsidation. It contains several inverted repeats capable of forming a secondary structure that can be recognized during the packaging of the pgRNA into the capsid. The 3.35 kb and 3.30 kb transcripts (table 1.1) both contain ε but only the shorter one is encapsidated. ε is highly conserved among the members of the family *Hepadnaviridae*. The functions of ε include, activation of the viral polymerase, template restriction, and initiation of minus strand DNA synthesis. Mutations within ε can greatly affect its function (Kramvis and Kew, 1998).

1.1.5.3. POLYADENYLATION SIGNAL

All HBV transcripts share a common polyadenylation signal located approximately 20 bp from the 3' end of the negative strand (figure 1.3) (Schaller and Fischer, 1991). The

polyadenylation signal is only recognized after its second encounter with the transcription complex (Russnak, 1991). Regulation of the polyadenylation signal involves a positive and a negative effector of transcription (*PET* and *NET*). *NET* enables the bypass of the transcription stop site during pregenome transcription, whereas *PET* enables termination of transcription on the second encounter (Beckel-Mitchener and Summers, 1997, Huang and Summers, 1994).

1.1.6. HBV LIFE CYCLE

Upon infection with HBV, the complete virion particle binds to the surface of the hepatocyte (Seeger and Mason, 2000). This binding takes place via the pre-S1 domain of the surface protein and is probably facilitated by a bridge of modified albumin (Machida *et al*., 1984, Neurath *et al*., 1986). Although a specific receptor for DHBV, the duck carboxypeptidase D (DCPD) have been identified, its human equivalent still remains elusive (De Falco *et al*., 2001, Kuroki *et al*., 1994, Kuroki *et al*., 1995, Seeger & Mason, 2000, Urban *et al*., 2000). Tissue specificity of HBV is regulated both at the level of tissue specific receptors and at the transcriptional level by liver specific transcription factors hepatocyte nuclear factor 1, 3 and 4 (HNF-1, HNF-3 and HNF-4) (Seeger and Mason, 2000, Sung and Lai, 2002). After penetration, the capsid is released into the cytoplasm where it is transported to the nucleus. In the nucleus the partially double-stranded DNA molecule is then released from the nucleocapsid. This is followed by its conversion to cccDNA (Ruiz-Opazo *et al*., 1982a, Ruiz-Opazo *et al*., 1982b). The cccDNA serves as a template for transcription of viral mRNAs, namely:

the precore mRNA, pgRNA (core mRNA), pre-S1 mRNA, pre-S2 mRNA, and X mRNA.

Viral replication begins with the binding of the viral reverse transcriptase to the 70-nucleotide stem-loop encapsidation signal (ε) at the 5' end of the pgRNA (Summers and Mason, 1982) (figure 1.5.) This is followed by initiation of reverse transcription by a protein priming mechanism (Wang and Seeger, 1992). Following synthesis of three bases complementary to the bulge of ε, the polymerase translocates to the 3'DR1 of the pgRNA. Here the bases bind to their complementary sequences and continue the synthesis of the negative strand (Wang and Seeger, 1993). The pgRNA template is degraded by RNase-H-like activity as DNA synthesis proceeds, except for the 17 or 18 bases that serve as a template for the plus stand synthesis upon translocation to DR2 (Nassal and Schaller, 1996). A final translocation, facilitated by the 9 base terminal redundancy, occurs when the plus strand reaches the 5' end of the minus strand allowing for circularization of the viral genome.

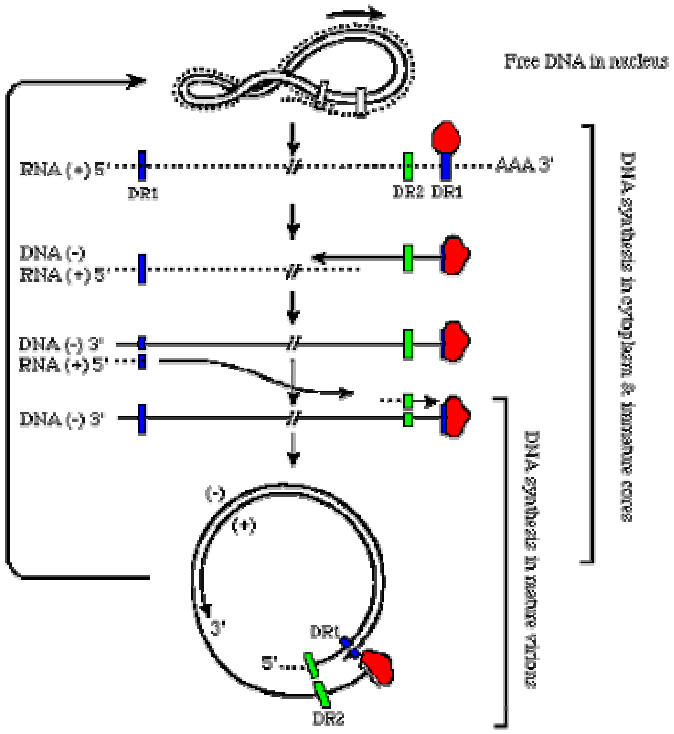


Figure 1.5. A diagrammatic representation of the replication process of HBV DNA. DR1 and DR2 depict direct repeat sequences, consisting of 12 bp each (http://www.stanford.edu/group/virus/1999/tchang/replication.htm).

A number of virions may have double stranded linear genomes as a result of lack of translocation from DR1 to DR2 (Staprans *et al*., 1991). After synthesis of the partially double-stranded DNA genome, nucleocapsid maturation occurs followed by coupling with envelope proteins in the endoplasmic reticulum (ER). The mRNA translation and assembly of some translated products into core particles takes place in the cytoplasm using the host cell machinery. The viral core can either be recycled back to the nucleus or be secreted as mature viral particles by interaction of the surface proteins and the endoplasmic reticulum.

Figure 1.6. HBV life cylce. The life cycle begins with the attachment of the virus to the as yet unidentified cellular receptors. It then uncoats releasing the nucleocapsid that migrates to the nucleus where, the partially double stranded DNA of the virus is converted into covalently close circular DNA (cccDNA). This serves as a template for transcription of viral mRNAs, which are subsequently translated in the cytoplasm. The 3.5 kb mRNA (pregenome) serves as the template for the synthesis of the core protein and reverse transcriptase (RT), the RT is then packaged into the core particles together with the pregenome, and this is followed by viral DNA synthesis by reverse transcription. The nucleocapsid can then be redirected to the nucleus to increase the cccDNA pool or be coated by surface proteins and released from the cell.

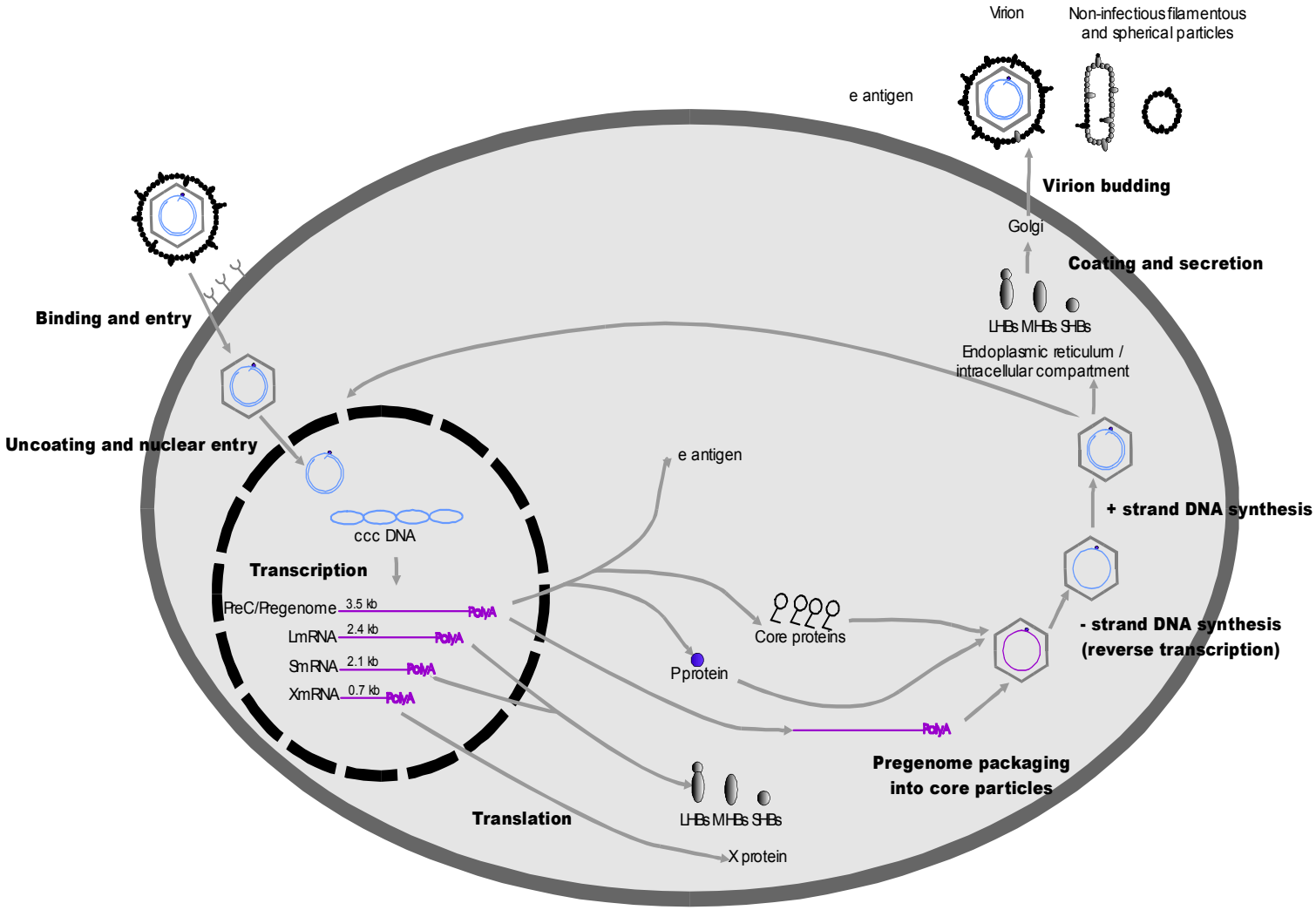

Virion
Non-infectious filamentous and spherical particles
e antigen
Virion budding
Golgi
Coating and secretion
LHBs MHBs SHBs
Endoplasmic reticulum / intracellular compartment
Binding and entry
Uncoating and nuclear entry
e antigen
+ strand DNA synthesis
ccc DNA
Transcription
PreC/Pregenome 3.5 kb
LmRNA 2.4 kb
SmRNA 2.1 kb
XmRNA 0.7 kb
PolyA
Core proteins
P protein
- strand DNA synthesis (reverse transcription)
Pregenome packaging into core particles
Translation
LHBs MHBs SHBs
X protein

Figure 1.6

1.1.7. IMMUNO-PATHOGENESIS OF HBV INFECTION

1.1.7.1. TRANSMISSION

HBV is transmitted by percutaneous or permucosal transmission of infectious body fluids. This may be achieved by exposure to contaminated blood or blood products by sexual contact, needle-stick injuries, the use of contaminated needles during intravenous drug abuse, other breaches in skin, via mucous membrane of gastro-intestinal tract, and perinatally from mother to child. In regions of low endemicity primary HBV infections occur mostly in adults in defined risk groups whose life-style places them at risk of infection. Transmission in this area is mostly percutaneous as a result of illicit use of intravenous drugs or by sexual contact in heterosexuals or homosexuals with multiple sex partners. In endemic areas, primary HBV infections occur mostly perinatally from infected mother to child or in early childhood through horizontal spread, namely, from recently infected and hence highly infectious young siblings or playmates (Ahn, 1996, Chiaramonte *et al*., 1991, Kew, 1996b, Maddrey, 2001, Mahoney, 1999).

1.1.7.2. PATHOGENESIS OF HBV INFECTION

Primary HBV infection is usually acute in adults and mild in children. The incubation period is six to twenty four weeks (Mahoney, 1999). The immune response initiated by cytotoxic and helper T-cell to viral antigens is of vital importance for viral clearance and disease pathogenesis in HBV infection. The incubation period influences the outcome of HBV infection because virus-specific immune responses and reduction in

viral replication occur during this period (Webster and Bertoletti, 2002, Webster *et al*., 2000).

Infection with HBV can occur in several patterns. Most primary infections in adults are self-limiting, clinically evident, and are usually resolved within 1-6 months, whereas those in young children are usually mild or sub-clinical and often result in chronic asymptomatic infection (Seeger and Mason, 2000).

1.1.7.3. SELF-LIMITED (ACUTE) HBV INFECTION

Acute HBV infection may produce severe liver disease that leads to fulminant hepatitis in approximately 1% of patients (**http://www.who.int**) (Webster and Bertoletti, 2002). Fulminant hepatitis B is caused by massive necrosis of the liver and is usually fatal. Ninety to ninety five percent of HBV infections in adults are acute (Webster and Bertoletti, 2002). After infection of hepatocytes by HBV the cytotoxic T lymphocytes and helper-T- cells rapidly recognize a number of different epitopes within the core, polymerase, and envelope proteins of the virus, presented to them by the major histocampartibilty complex type I and II on the surface of the infected cells. This leads to necrosis of the infected cells or secretion of antiviral cytokines that interrupt the HBV life cycle within the infected cells (Jung and Pape, 2002). Prevention of the virus from spreading can be achieved by production of viral antibodies as well as local interferons.

Acute Hepatitis B infection usually lasts between 1-6 months, and includes an asymptomatic (latent) stage characterized by high levels of viraemia. There is

increasing evidence that maximal reduction in HBV-DNA concentration, together with an increase in cytokine mRNA and antigen specific CD-8 cells occurs at this stage. Early suppression of the immune system by the virus may be achieved with the aid of HBeAg, which acts as a decoy (Chen *et al*., 2004) and the ability of HBV to inhibit IFN-β and block the activation of INF-α pathways (Milich *et al*., 1998, Whitten *et al*., 1991). After the latent period of infection there is concurrent clearance of infected cells from the liver as well as the virus from serum (Seeger and Mason, 2000).

Treatment of acute hepatitis B should theoretically shorten the clinical course, accelerate recovery and eliminate viral persistence. Unfortunately, all antiviral drugs to date (i.e interferon-alpha, lamivudine and famciclovir) have been found to be largely ineffective in the treatment of acute HBV infections (Akuta and Kumada, 2005).

1.1.7.4. PERSISTENT (ASYMPTOMATIC) HBV INFECTION

Patients who remain HBsAg positive for more than six months after HBV infection almost always become chronically infected. This is frequently asymptomatic. Chronic HBV infection is defined as the presence of HBsAg in the serum of a healthy individual for 6 months or longer (Seeger and Mason, 2000). When asymptomatic these individuals have normal serum amino tranferase levels. This infection may persist for life and may ultimately lead to cirrhosis and/or hepatocellular carcinorma (HCC) (Feitelson, 1994).

Asymptomatic infection occurs when the immune system fails to completely clear HBV infection. The frequency of viral persistence following acute infection is related to age and immune deficiency: up to 10% in adults, 30% in children age 1-5, 90% in neonates (McMahon *et al*., 1985). Immune-deficient individuals such as those with renal insufficiency requiring hemodialysis (Szmuness *et al*., 1974), and those with Down's syndrome (Hollinger *et al*., 1972) also have a high risk of becoming chronic carriers. Clearance of HBV in chronic infections may occur spontaneously, with seroconversion to negativity for HBeAg and HBsAg in 10% and 1 to 2% of cases per year, respectively (McMahon *et al*., 1985). Defining asymptomatic HBV infection only by use of the HBsAg positivity is limiting, because in some cases where the integrated virus produces an intact HBsAg, this can be confused with the presence of complete viral particles. In addition, HBV DNA may be detected in the serum and liver tissue of individuals in whom HBsAg can no longer be detected by serological assays (Owiredu *et al*., 2001a). Five to ten percent of asymptomatic HBV infected patients have high levels of HBV DNA (Bonino *et al*., 1986). HBeAg can be detected in 5-50% of all persistently infected individuals. Persistent infection without detectable levels of HBsAg has been documented. This might be the result of production of very low levels of HBsAg, or mutations in the HBsAg epitope targeted by the serological test (Allain, 2004).

Treatment of chronic HBV should theoretically eliminate viral replication and prevent progression to cirrhosis and HCC. To date interferon-alpha has been the most effective treatment of chronic HBV infection (Wong *et al*., 1993).

1.1.7.5. HEPATOCELLULAR CARCINOMA (HCC)

HBV infection is incriminated in the aetiology of about 80% of HCC cases worldwide (Kew, 2002). About 330,000 people develop HBV-related HCC each year in sub-Saharan Africa and the Far East (Bosch and Ribes, 2000, Bosch *et al*., 1999).

The global geographic prevalence of chronic HBV infection and HCC correspond (Beasley, 1988, Beasley *et al*., 1981, Kew, 2002): south east Asia, some of the Pacific islands and sub-Saharan Africa have HBV carrier rates of 7-15% and the highest incidences of HCC, whereas countries with low carrier rates of the virus have low incidences of HCC (Kew, 2002, Simonetti *et al*., 1991).

HBV may be directly or indirectly involved in the pathogenesis of HCC. The direct contribution includes integration of the viral genome into the chromosomes of the infected hepatocytes (Kew, 1996a). Integration may occur by a mechanism of illegitimate recombination (Moradpour and Wands, 1994, Yang and Summers, 1995, Yang and Summers, 1999). The integrated viral DNA is usually extensively deleted, mutated, or rearranged (Moradpour and Wands, 1994). Unlike in woodchucks, human HBV integration does not usually occur at or near the site a proto-oncogene and this has led to the investigation of other possible mechanism by which HBV can directly cause HCC (Transy *et al*., 1992).

As the search for common oncogene targets of HBV continues to be unproductive, alternative mechanisms for oncogenesis have been investigated. One of these has been

the possible action of the HBx (X gene product) as a transcriptional transactivator (Colgrove *et al*., 1989, Twu and Schloemer, 1987, Zahm *et al*., 1988). HBx does not bind any known cellular genes; hence its transactivating function is most probably through its effect on cellular transcription factors (Seto *et al*., 1990). The function of HBx as a transactivator has been strengthened by the fact that 90% of transgenic mice expressing HBx developed HCC, and that immunohistochemistry of the liver tissue of these animals revealed high levels of HBx (Kim *et al*., 1991). The figure below summarizes the spectrum of liver diseases after HBV infection.

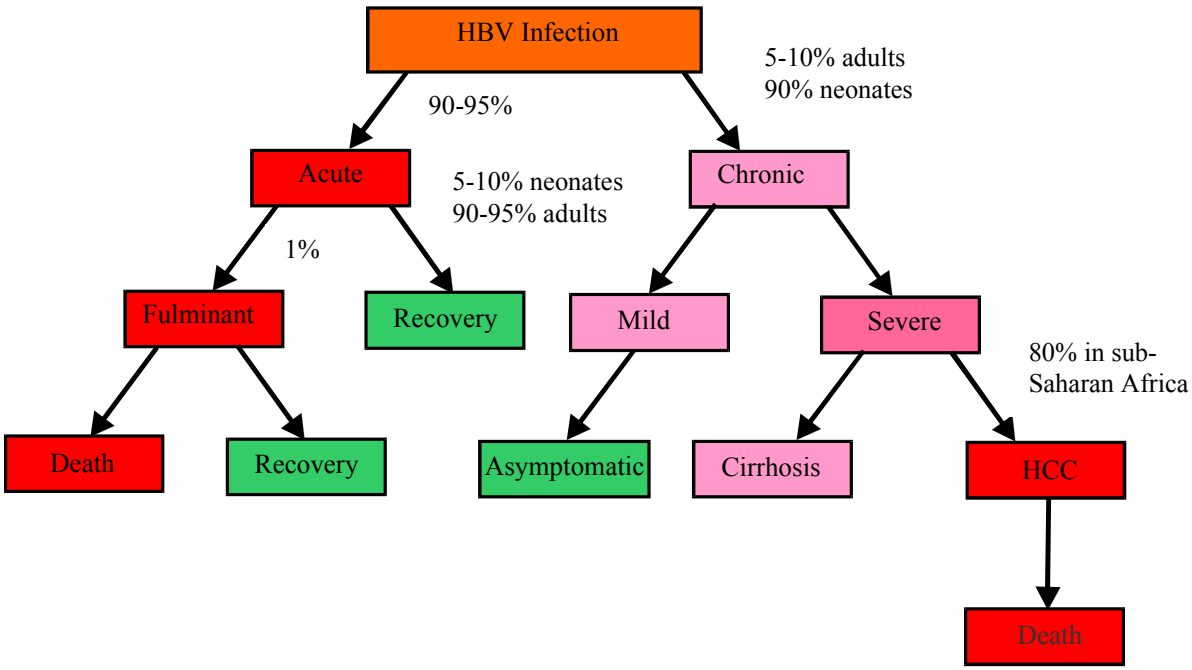


Figure 1.7. Spectrum of liver diseases after HBV infection. The infecting dose of HBV and the age of the person infected are important factors that correlate with the pathogenesis observed.

1.1.8. HBV GENOTYPES: MOLECULAR CHARACTERISTICS AND GEOGRAPHIC DISTRIBUTION.

A genotype is defined as the genetic constitution of an organism. In viruses it represent the most stable and replication competent genomes (Kramvis *et al*., 2005). Upon phylogenetic analysis HBV separates into 8 genotypes (genotypes A-H) based on complete genome divergence in the region of 8% (Norder *et al*., 1993, Stuyver *et al*., 2000) and more than 4% at the level of the S gene (Norder *et al*., 1992a) (Table 1.2).

Table 1.2. Distrubution of serological subtypes and genotypes of HBV.

Genotype	Serotype	Area of predominance
A	*adw2* and *ayw1*	Northwestern Europe, United States of America, Brazil and southern and central Africa
B	*adw2* and *ayw1*	Taiwan, Japan, Indonesia, China, Vietnam, and Oceania
C	*adw2, adrq+, adrq-, ayr*	East Asia, Taiwan, China, Korea, Japan, Polynesia, Oceania, Aborigines (Australia) and Vietnam.
D	*ayw2, ayw3* and *adw3*	Mediterranean area, India, north Africa, southern Africa and the middle east.
E	*ayw4*	West Africa
F	*adw4q-, adw2, and ayw4*	Central and south America, Polynesia
G	*adw2*	France, Germany, United States of America, and the United Kingdom
H	*adw4*	Central America

Data compiled from: (Bowyer *et al*., 1997, Kao, 2002, Kao and Chen, 2002, Magnius & Norder, 1995, Miyakawa & Mizokami, 2003, Odemuyiwa *et al*., 2001, Owiredu *et al*., 2001a, Owiredu *et al*., 2001b, Stuyver *et al*., 2000), Vieth *et al*., 2002 and (Mulders *et al*., 2004). Note that subtype *adw2* is the most widely distributed.

HBV can be divided into four major serological subtypes *adw, adr, ayw, ayr*. These can be further divided into ten subtypes according to the antigenic determinants of the surface antigen based on the deduced amino acid composition at positions 122, 127, and 160. The subtypes are: *ayw*1, *ayw*2, *ayw*3, *ayw*4, *ayr*, *adw*2, *adw3*, *adw*4, *adrq*$^{+}$ and *adrq*$^{-}$ (Arauz-Ruiz *et al*., 2002, Magnius and Norder, 1995, Norder *et al*., 1992) (Table 1.3).

Table 1.3. The serological subtypes of HBV.

Position / Subtype	aa residues specifying serological subtypes		
	122	127	160
ayw1	Arg	Pro	Lys
ayw2	Arg	Pro	Lys
ayw3	Arg	Thr	Lys
ayw4	Arg	Leu	Lys
Ayr	Arg	Pro	Arg
adw2	Lys	Pro	Lys
adw3	Lys	Thr	Lys
adw4	Lys	Leu	Lys
*adrq-**	Lys	Pro	Arg
*adrq+**	Lys	Pro	Arg

**adrq-* differs from *adrq+* at residues 158 and 159, and at residues 177 and 178 with *adrq-* subtypes having amino acids substitutions in nearby positions.

The HBV genotypes appear to have distinct geographic distributions. Genotypes A prevail in Northwestern Europe, United States of America, Brazil, southern Africa and central Africa. Genotype D prevails in the Mediterranean area, India, north Africa, southern Africa and the middle East. Recently genotype A has also been found in west Africa (Cameroon) (Mulders *et al*., 2004).Genotype B and C are predominant in Asia and the Far East. Genotype E is restricted to Africa, predominating in west and central Africa. Genotype F is prevalent in Central America. The exact distribution of G has not been ascertained because of the small number of genomes sequenced from this group, which has been detected in from France, Germany, United Kingdom and United States of America. The first sequenced strains of genotype H have been from Central America and California. They are more closely related to genotype F than the other genotypes.

A closer look at the distribution of HBV in Africa, reveals the predominance of genotype D in North Africa; genotype A in southern Africa, parts of west and east Africa and genotype E in west and Central Africa (figure 1.8). Genotypes A, B, C, and D have been shown to be circulating in South Africa with A and D predominating. Genotype B and C have probably been introduced to South Africa by immigrants from South East Asia.

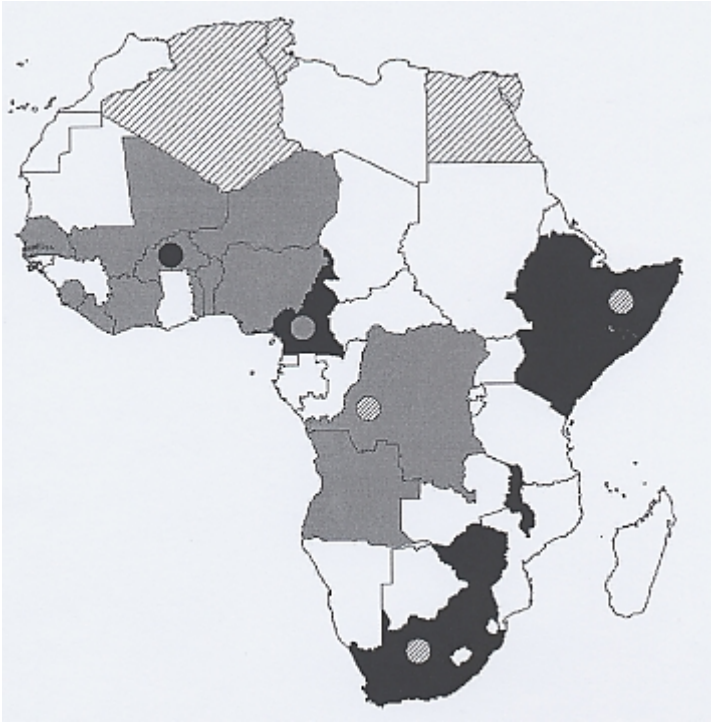


Figure 1.8. HBV genotype distribution across Africa. HBV genotype A, black; HBV genotype D, stripes; and HBV genotype E, gray. Minority genotypes are shown as inserted circles (Mulders *et al*., 2004).

1.1.8.1. GENOTYPES AND SUBGENOTYPES OF HBV

HBV genomes can be classified into subgenotypes when the nucleotide divergence between members of the same genotype is in the region of 4% at the level of the complete genome (Kramvis *et al*., 2005). So far subgenotypes of HBV have been identified in five genotypes i.e A, B, C, D and F.

Genotype A

Genotype A has a unique 6-nucleotide insert at the carboxy terminus of the core gene. Genotype A was the first genotype in which subgenotypes were recognized and named subgroup A' and subgroup A-A' (Bowyer *et al*., 1997). At that time, subgroup A' was thought to be present in South African isolates only. Subsequent studies, however have identified subgroup A' in isolates from Malawi (Sugauchi *et al*., 2003), Philippines, India, Nepal, Bangladesh (Sugauchi *et al*., 2004a), and most recently Brazil (Araujo *et al*., 2004). In the 2004 paper by Sugauchi *et al* subgroup A' was renamed as subgroup Aa (Genotype **A** from **A**frica and **A**sia). Subgroup A-A' was also called subgroup Ae (Genotype **A** from **E**urope) (Sugauchi *et al*., 2004a). A third subgenotype of genotype A which differs from subgroup Aa and subgroup Ae called subgroup A'', has been identified in Cameroon (Mulders *et al*., 2004). Subgroup A'' isolates were mostly of the *ayw1/2* subtype, unlike the Ae and Aa isolates which are mostly of the *adw2* subtype (Mulders *et al*., 2004, Sugauchi *et al*., 2004a).

Nomenclature of subgenotypes (previously called subgroups) according to geographic regions (Sugauchi *et al*., 2004a) is confusing because different subgenotypes of the

same genotype may be found in the same region e.g subgroup Aa and subgroup A'' are both found in Africa (Sugauchi *et al*., 2004a and Mulders *et al*., 2004). Moreover, new isolates of the already defined subgenotypes may be found in other regions of the world as has been found for subgroup Aa (Araujo *et al*., 2004). Therefore, it has been recommended that subgenotypes be named numerically (Kramvis *et al*., 2005). Hence subgroup Aa has been renamed subgenotype A1, whereas subgroup Ae is called subgenotype A2. This nomenclature will be used in this thesis.

Genotype B

Genotype B was originally classified into two subgenotypes based on recombination with genotype C in the precore region between 1740-1838 and 2443-2485. Strains without recombination all came from Japan and were classified as subgroup Bj (subgenotype B1), whereas those with recombination came from other Asian countries (China, Hong-Kong, Indonesia, Japan, Taiwan, Thailand, and Vietnam) and were classified into subgroup Ba (subgenotype B2) (Sugauchi *et al*., 2004b, Sugauchi *et al*., 2002).

Subsequent analysis of a larger number of HBV strains belonging to genotype B has classified this genotype into four subgenotypes B1 to B4 (Norder *et al*., 2004). Subgenotype B1 consists mainly of strains previously classified as subgroup Bj. Subgenotype B2 (formerly called subgroup Ba) comprise mainly of genotype B strains from China, Taiwan, and Hong Kong and to a lesser extent Vietnam and Thailand. Subgenotype B3 was made up of four isolates from Indonesia. Subgenotype B4

comprised mainly of strains from Vietnam. Subgenotype B1, B2 and B3 specified *adw2*, whereas subgenotype B4 isolates specified *ayw1* (Norder *et al*., 2004).

Genotype C

Genotype C in earlier studies was classified into two major subgenotypes, subgroup C1 and subgroup C2. Subgroup C1 isolates were mostly from south east Asia (Vietnam, Myanmar and Thailand and southern China), whereas subgroup C2 isolates originated from far east Asia including Japan, Korea and northern China (Huy *et al*., 2004, and Henry *et al*., 2005). Serological subtypes *adrq+, adrq-, ayr* are specified only by genotype C isolates.

Analysis of 66 complete genotypes C genomes by Norder et al have classified genotype C into four subgenotypes C1 to C4 (Norder *et al*., 2004). The separation of genotype C into four subgenotypes is also evident in the study by Chan *et al* (Chan *et al*., 2005). Subgenotype C1 (subgroup C2 or genotype Ce) isolates were mainly from the Far East, whereas subgenotype C2 (subgroup C1 or genotype Cs) isolates were from south east Asia. Two branches formed subgenotype C3. One branch comprised strains from aboriginal populations from New Zealand to Polynesia, whereas the other branch consists of strains from Micronesia and Japan. Subgenotype C4 was the most diverse and specified *ayw3*. Strains belonging to subgenotype C4 were isolated from Aborigines in northeast Australia.

Genotype D

Genotype D is characterized by a unique 33 bp nucleotide deletion at the beginning of the preS1 gene. It is the most widely distributed genotype geographically. Genotype D can be divided into four subgenotypes D1 to D4 (Norder *et al*., 2004). The distribution of subgenotypes of genotypes D is not clearly demarcated but a pattern can be elucidated. Subgenotype D1 consists mainly of isolates from the Middle East, subgenotype D2 isolates were mainly from India and Germany. Subgenotype D3 isolates were mainly from Alaska, Sweden, France and South Africa, whereas D4 isolates were mainly from Somalia and Oceania.

Genotype E

Genotype E is very closely related to genotype D. There is no separation between these two genotypes in the C and X ORFs (Bowyer and Sim, 2000). Genotype E has a 3-nucleotide deletion at the amino terminal of the preS1 region (Norder *et al*., 1994), which it shares with genotype G. An A3095G mutation within the D region of the S2 promoter (Moolla *et al*., 2002) is unique to genotype E isolates. This nucloetide substitution introduces a new translational start codon downstream from the pre-S1 start codon. This may result in the translation of an elongated MHBs protein (317 aa in length instead of 281 aa) (Kramvis *et al*, 2005 in press). Moreover, though the initiation context of the introduced AUG is weak, it may also interfere with translation of the MHBs by means of a leaky scanning mechanism (Kozak, 1997). Low divergence between genotype E members (Kramvis *et al*, 2005 in press) suggests a recent introduction of this genotype into human population. However, some studies have

found genotype E in African-Americans in Northern America (Kato *et al*., 2004, Westland *et al*., 2003) suggesting a longer existence.

Genotype F

The relatively high nucleotide divergence between genotype F and the remainder of the HBV genotypes suggest that genotype F may represent the first split from the human hepadnaviral ancestor (Norder *et al*., 1994). Genotype F can be divided into two subgenotypes based on a nucleotide substitution at position 1858 of the S gene. Subgenotype F1 (T1858 Thr^{45}) represents genotype F isolates from Central America while subgenotype F2 (C1858 Leu^{45}) represents those from South America and Polynesia (Norder *et al*., 2003).

Genotype G

All genotype G isolates possess a 36 bp insert at position 1905 within the core gene. The precore region process two translational stop codons (positions 2 and 28) abrogating the production of the HBeAg (Kato *et al*., 2002). However, the HBeAg has been detected in sera of patients infected with genotype G. This can be explained by the fact that infection with genotype G is usually together with a wild-type virus belonging to another genotype (Kato *et al*., 2001, Kato *et al*., 2002).

Genotype H

HBV genotype H is very similar to genotype F, and may have split off from genotype F within the New World (Arauz-Ruiz *et al*., 2002). It can be differentiated from genotype by the presence of two unique mutations on the HBsAg (Val^{44} and Pro^{45}).

The distribution of HBV into several genotypes result from a high genetic variability of HBV isolates around the world. This variability results from the pressure of the host on the virus and the relatively high error rate of the viral reverse transcriptase.

1.1.9. HBV MUTANTS: DISTRIBUTION AMONG GENOTYPES AND INFLUENCE ON OUTCOME OF HBV INFECTION.

Considering the limitations imposed on mutations by the overlapping reading frames (Mizokami *et al*., 1997), the mutation rate of the hepadnaviruses is relatively high (1.4×10^{-5} to 5×10^{-5} nucleotide substitutions per site per year) (Kramvis *et al*., 2005). This is the result of the relatively high viral turnover during infection, the lack of proofreading activity of the viral reverse transcriptase, pressure of the host immune response and antiviral treatments.

1.1.9.1. PRECORE/ CORE GENE MUTANTS

Because a large region of the precore ORF is not overlapped by other ORFs, more mutations are tolerated in this section. Most of the precore/core mutations cluster in the middle third of the core gene. This region harbours the B and T cell epitopes, although none of the mutations have been proven to result in loss of immune recognition (Pumpens and Grens, 2001).

One of the most important mutations in the precore is the G1896A point mutation. This mutation transforms codon 28 of the precore gene from TGG to a TAG, creating a stop codon within the precore gene, and resulting in a 28aa truncated HBeAg (Carman *et al*., 1989). The G1896A mutation occurs within the encapsidation signal and is favoured by a T1858. Base pairing with T1858 results in better viral genome stability, pregenomic encapsidation, and initiation of DNA synthesis (Lok *et al*., 1994). This mutation does not affect packaging of the pregenome because the core protein and HBeAg are translated from different mRNAs (Ngui *et al*., 1999). Because the presence of the A1896 depends on T1858 or C1858, it shows a geographic distribution corresponding to that of the genotypes (Li *et al*., 1993). Genotype B, D, and E have T1858, whereas A and H have C1858. Genotype C and F isolates have either T1858 or C1858. In genomes with T1858, a Watson-Crick base pair with A1896 stabilizes the mutant A1896 and this in turn stabilizes the encapsidation signal (Rodriguez-Frias *et al*., 1995). All genotype A, and genotype H isolates and some of genotype C have a C at 1858 and the G→A change at 1896 produces a 'wobble pair' resulting in destabilization of the stem-loop structure and impairment of its functionality (Li *et al*., 1993, Rodriguez-Frias *et al*., 1995). Therefore the 1896 mutation is uncommon in genotypes that have C1858.

Although initially associated with severe liver disease, including liver failure (Omata *et al*., 1991), the 1896A mutant has also been identified in chronic HBV carriers (Lindh *et al*., 1996, Minuk *et al*., 2000). Moreover, later studies indicate that reactivation of disease is not associated with the 1896A mutant (Loriot *et al*., 1995). This, together with the observation that HBV-DNA levels are not elevated in patients with precore

1896A mutant infections, suggest that co-mutations might explain the virulent forms of pre-core infections originally associated with this mutant (Knoll *et al*., 1999).

Another set of mutations that occurs within the precore gene is the C1856T and G1862T. These mutations result in an amino acid change from proline to serine in codon 15 and valine to phenylalanine in codon 17 (Kramvis and Kew, 1999, Baptista *et al*., 1999, and Lok *et al*., 1995). The valine to phenylalanine substitution occurs close to the signal peptide site, which lies between codon 18 and 19. The presence of an aromatic residue like phenylalanine at the –3 position of the signal peptidase recognition motif is not tolerated (Von Heijne, 1983, and Von Heijne, 1984), as it may prevent removal of the 19 aa of the precore/core fusion protein during maturation of HBeAg. C1856T is predominant in genotype D, B, C, E, and G. G1862T has been associated with fulminant HBV infection in genotype B. This is caused by reducing HBeAg production from its precore protein precursor (Hou *et al*., 2002).

The core accounts for the highest number of deletion mutants reported in the HBV genome. These deletion mutants survive through a defective interference-like mechanism because they usually co-exist with wild type viruses (Marinos *et al*., 1996, Yuan *et al*., 1998a, Yuan *et al*., 1998b). They may also be involved in immune escape, because deletion usually involves the B and T cell epitopes. Some studies also indicate that HBV variants with core gene deletions may inhibit HBV replication (Marinos *et al*., 1996). The core insertions are less common than deletions and have not been

associated with any process in the life cycle of the virus or any change in pathogenesis (Ngui *et al*., 1999).

1.1.9.2. BASIC CORE PROMOTER (BCP) MUTANTS

The BCP drives the synthesis of the 3.5 kb pregenome and precore transcripts. The most important mutations in the BCP are the A1762T and G1764A mutations (Okamoto *et al*., 1994). Viruses with these mutations may or may not posses the 1896 mutation. These mutations are found in all genotypes and represent the 'wild-type' genome in genotype G (Stuyver *et al*., 2000).

These mutations are frequently reported in chronic HBV infected patients, liver transplant patients, and HCC patients, but rarely in asymptomatic HBV carriers (Baptista *et al*., 1999, Blackberg and Kidd-Ljunggren, 2003). These mutations reduce the amount of HBeAg being produced by reducing the amount of precore mRNA because they occur within the binding site of the liver transcription factors (Buckwold *et al*., 1996, Gunther *et al*., 1998a). There have been conflicting reports regarding the effect that these mutations have on the replication of the virus. This might be related to genotypes in which these mutants were present (Buckwold *et al*., 1996, Gunther *et al*., 1998a) .

Insertions and deletions have also been reported in the BCP (Kramvis and Kew, 1999). Insertion within this region, which usually results in an increase in transcription binding sites may occur as a compensatory mechanism for core gene deletions that have a

negative effect on core gene synthesis (Gunther *et al*., 1996a). Deletions within this region frequently result in enhanced production of replicative intermediates and increased secretion of DNA containing particles (Gunther *et al*., 1996b).

1.1.9.3. X GENE MUTANTS

HBx functions in stimulation of viral gene expression and replication, hence infection with X gene mutants is characterized by low levels of viraemia (Feitelson, 1994). Some data suggest that HBx may inhibit clonal outgrowth of cells and induce apoptosis by a p53-independent pathway (Sirma *et al*., 1999). If HBx mutants are selectively expressed in HCC patients (Poussin *et al*., 1999), they may enhance hepato-carcinogesis by abrogating the anti-proliferative and apoptotic effects of HBx.

Infection with X gene deletion mutants may occur independently of wild-type viruses (Feitelson, 1994). The X gene deletion mutants most often reported are those containing an 8 nucleotide or a 20 nucleotide deletion at the 3'end of the X gene (Fukuda *et al*., 1996, Laskus *et al*., 1994, Repp *et al*., 1992, Uchida *et al*., 1995). These deletions may result in a decrease HBsAg, HBeAg, and HBcAg titers (Moriyama, 1997).

1.1.9.4. S GENE MUTANTS

The S gene is vital in the life cycle, as well as infectivity of the HBV virus. The target of the host's humoral response to HBV is the hydrophilic region of the HBsAg between aa residues 100 and 160. The most important mutants of the S antigen are the vaccine

escape mutants (Carman *et al*., 1990). These mutations may result in escape from the host immune response, resulting in acute infection in individuals immunized against HBV. In these individuals anti-HBs and HBsAg co-exist. The prototype of the vaccine escape mutants is the Gly→Arg mutation at aa 145 of the S gene (Carman *et al*., 1990) (figure 1.9). Other common mutations that have been described in this region include Thr126→Asn, Gln129→His, Met133→Leu, and Asp144→Ala (Hunt *et al*., 2000) (figure 1.9).

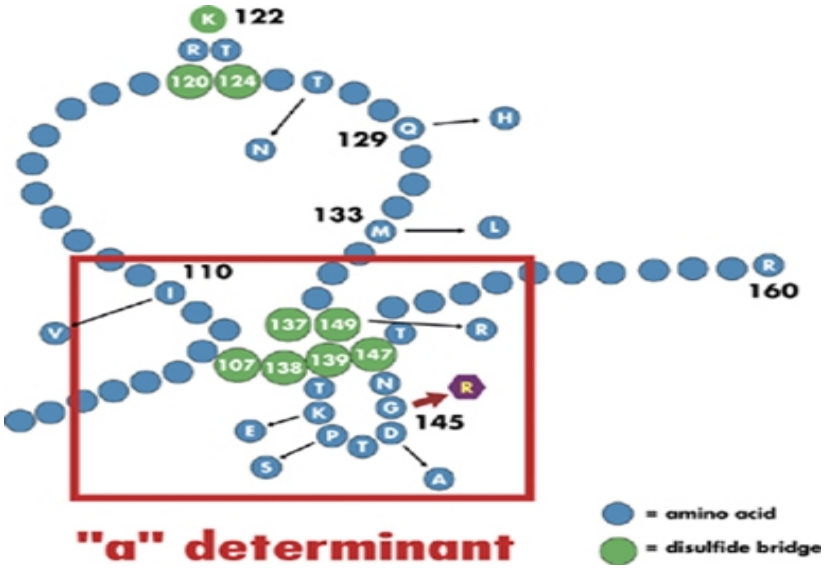


Figure 1.9. 'a' determinant of HBsAg showing the prototype vaccine escape mutant in purple. Other reported vaccine escape mutants are indicated. Disulphide bridges are formed between aa in green.

Nucleotide insertions have been reported within aa 121-124 of the HBsAg. Insertions within this region may result in false-negative HBsAg testing and therefore represent a risk of misdiagnosis to the safety of products used for blood transfusion (Yamamoto *et*

al., 1994, Carman *et al*., 1995, and Hou *et al*., 1995). A 183nt deletion involving the S promoter is the most common deletion of the S gene (Cabrerizo *et al*., 2000, Gerner *et al*., 1998). Mutants with this deletion co-exist with wild-type genomes *in vivo* although they are capable of infecting, replicating and releasing viral particle *in vitro* (Cabrerizo *et al*., 2002). Pre-S2 deletion mutants are fully functional. They may or may not co-exist with wild-type viruses (Fernholz *et al*., 1993, Gerner *et al*., 1998, Kidd-Ljunggren *et al*., 1995, Mulders *et al*., 2004).

1.1.9.5. POLYMERASE GENE MUTANTS

Natural mutations within the polymerase gene are expected to be lethal because the protein encoded by this gene is responsible for DNA polymerase, RT and RNase H activity, as well as packaging of the pregenome, all vital steps in the HBV life cycle.

Mutations within the polymerase gene are particularly common during HBV therapy with nucleoside analogues, which act by inhibiting the function of viral DNA polymerase such as lamivudine (Papatheodoridis *et al*., 2002 and Severini *et al*., 1995). These mutations usually occur within the major catalytic region of the viral polymerase containing the Tyr-551-Met-552-Asp-553-Asp-554 (YMDD) motif. This region is conserved in all reverse transcriptases and overlaps the ‘a’ determinant (Oon *et al*., 1999). The most commonly reported mutations are the substitution of Met-552 with Val or Ile (M552V or M552I). Another mutation resulting in the substitution of Leu at position 528 with Met is also common (Papatheodoridis *et al*., 2002). The replication rate of these mutants (M552V and M552I) is very low when compared with that of wild-type viruses, explaining the reappearance of wild-type viruses after termination of

lamivudine therapy (Chayama *et al*., 1998, Melegari *et al*., 1998, Ono-Nita *et al*., 1999a, Ono-Nita *et al*., 1999b).

There is a correlation between some HBV mutations and clinical outcome of infection. Because individual mutations can be associated with various genotypes, genotypes are expected to have a considerable influence on the outcome of HBV infection.

1.1.10. GENOTYPE AND OUTCOME OF HBV INFECTION

The clinical outcome of HBV infection may vary depending on the following: the age of the individual, initial immune response, and the virus strain infecting the individual. Interest in the influence of genotypes on the outcome of HBV infection has grown as the influence of different mutations on the clinical outcome of HBV infection became more evident. Comparative studies of different genotypes are restricted by the predominance of 1 or 2 HBV genotypes in most geographic regions. The exact reason why HBV genotype may be related to clinical outcomes is not clear. It is likely that different genotypes could be associated with differences in replication levels and expression of immune epitopes.

In one study it was shown that there was an association between genotype A and chronic infection (Mayerat *et al*., 1999). This could be the result a lower antigenicity of genotype A as compared with genotype D. Grandjacques and coworkers also found that genotype A was predominant in chronic infection in France when compared with genotype D (Grandjacques *et al*., 2000). They found that there were more anti-HBe

positive patients infected with genotype D than A, whereas HBeAg positive patients were more often genotype A. Genotype D has also been implicated in severe liver disease and development of HCC in Indian patients (Thakur *et al*., 2002).

The existence of genotype B and C in Asian populations in equal proportions has allowed for comparison of the relationship of genotypes to disease outcome. In one such study carried out in Japan, it was shown that HBeAg positivity was more frequent in genotype C than in B, with the reverse being true for anti-HBe. Secondly, the A1762T/G1764A, which has been shown to decrease HBeAg production (Buckwold *et al*., 1996), was shown to predominate in genotype C (Orito *et al*., 2001). The latter was confirmed by a study carried out by Kao *et al* in Taiwan (Kao *et al*., 2000a). Other studies in Japan and Tawain have also implicated genotype C with more severe liver disease, poor prognosis and development of HCC (Kao *et al*., 2002, Yuen *et al*., 2004a, Yuen *et al*., 2004b). In Thailand Bj has been postulated to be related to more severe liver disease and non-progression to HCC, whereas Ba is associated with development of HCC at a young age (Sugauchi *et al*., 2002).

Limited data is available on the relationship of subgenotypes and outcome of HBV infection. In a case control study carried out on southern African, individuals infected with subgenotype A1 were 4 times more likely to develop HCC than individuals infected with subgenotype A2 or other non A genotypes. One hundred percent of those infected with genotype A were infected with subgenotype A1 suggesting a higher hepatocarcinogenic potential of this strain (Kew *et al*., 2005).

Evidence indicates that HBV genotypes are as important in influencing the natural history of HBV infection as they are in influencing the response to antiviral therapy, in particular interferon. The antiviral effects of interferons are exerted through three main pathways:

- Increased expression of Class I and Class II MHC glyoproteins, thereby facilitating the recognition of viral antigens by the host immune system
- Activation of cells with the ability to destroy virus-infected targets
- Direct inhibition of viral replication.

(www.virology-online.com)

Kao *et al* in 2000 reported that genotype B infected patients were more responsive to interferon therapy than genotype C infected patients (Kao *et al*., 2000b). Genotype B was also found to be more responsive to lamivudine treatment than genotype C in Taiwan, although both groups of patients had equal risk of developing lamivudine resistance after a year of therapy (Kao, 2002). In a study on liver transplant patients, it was shown that patients with genotype A had a lower risk of HBV recurrence despite having higher levels of pre-transplantation HBV replication. In contrast, genotype D infected individuals appeared to have a higher risk of HBV recurrence and mortality (Devarbhavi *et al*., 2002).

Although all the above studies provide evidence that the outcome of HBV infection can be influenced by the genotype of HBV infecting the patient, further studies in other parts of the world, together with large scale longitudinal studies, followed by the

elucidation of the mechanism by which the different genotypes influence the outcome of HBV infection will be needed, to confirm these findings.

There is a clear indication that the outcome of HBV infection is genotype dependent. HBV genotype distribution varies geographically. Hence, studies based on genotypes from other regions of the world cannot be extrapolated to southern Africa. At the beginning of this study there was limited data on complete HBV genomes from south Africa and data suggesting unique mutations to subgebotypes from this region was base on subgenomic amplification of the HBV genome. The aim of this thesis was to augment the database of complete HBV genomes from southern Africa, characterize these complete genomes at a molecular level and functionally characterize some of the mutations identified.

1.2. AIMS AND ORGANIZATION OF THE STUDY

The aims of this study were to:

- **Extract HBV DNA from sera of asymptomatic carriers of HBV and from patients with acute hepatitis B from South Africa.**
- **Amplify the full genome of HBV isolates from asymptomatic carriers of the virus and from acute hepatitis B patients.**
- **Characterize the amplicons at molecular level.**
- **Functionally characterize some of the mutations identified.**

To achieve the aims above, the study was divided into 4 sections, which are presented in:

Chapter 2

1. DISTINCTIVE SEQUENCE CHARACTERISTICS OF SUB-GENOTYPE A1 ISOLATES OF HBV FROM SOUTH AFRICA. This study involved the amplification and molecular characterization of complete HBV subgenotype A1 isolates from South Africa. Although subgenotype A1 had already been identified in South Africa, the additional complete genomes allowed the identification of amino acids in the 4 ORFs and nucleotide substitutions in the *cis*-regulatory region that distinguish to subgenotype A1 from other genotypes and subgenotypes.

Chapter 3

2. EFFECT OF THE G1888A MUTATION OF SUBGENOTYPE A1 ON TRANSLATION OF THE CORE PROTEIN. Molecular characterization of subgenotype A1 isolates from South Africa lead to the identification of several unique nucleotide substitutions within the *cis*-acting elements. The G1888A substitution introduces an out-of-frame AUG, creating an overlapping uORF, which terminates five nucleotides downstream from the core AUG. In addition to stabilizing the encapsidation signal by forming a Watson and Crick base pair with T1871, this mutation may affect translation of the core protein, most probably by a leaky scanning mechanism. The aim of this study was to determine whether this mutation had any modulating effect on the translation of the core protein, and if, so the mechanism involved in this modulating role.

Chapter 4

3. ANALYSIS OF GENOTYPE D ISOLATES FROM AFRICA. Prior to this study there were no complete genome sequences of HBV genotype D from South Africa. This section involves the molecular characterization and phylogenetic analysis of six complete genotype D genomes from South Africa.

Chapter 5

4. During amplification of full genomes by a method involving two overlapping HBV genome fragments, a virus host junction of HBV DNA integrated into chromosome 7 in the WBSCR 1 gene was fortuitously amplified in one of the acute HBV infected

patients. The characterization of this integrant is described in this chapter and the implication of integration during acute HBV infection on the development of HCC is also discussed.

Chapters 2 and 4 have been introduced in chapter 1. The materials and methods for chapter 4 are the same as those in chapter 2.

CHAPTER 2

DISTINCTIVE SEQUENCE CHARACTERISTICS OF SUBGENOPTYPE A1 ISOLATES OF HBV FROM SOUTH AFRICA

SUMMARY

HBV can be divided into 8 genotypes based on genome divergence of about 8% according to phylogenetic relatedness. Genotypes A and D predominate in South Africa. Genotype A is the most frequent and has recently been shown to consist of two sub-genotypes A1 and A2, with A1 predominating. The complete genome of HBV isolates from 18 asymptomatic carriers and 5 patients with acute HBV B was amplified and the resulting amplicons were cloned and sequenced. No differences between isolates from asymptomatic carriers and those from patients with acute hepatitis B were found. All acute hepatitis B isolates belonged to subgenotype A1, ten isolates from asymptomatic individuals belonged to subgenotype A1, two to sub-genotype A2 and the remaining six to genotype D. The distinguishing amino acid residues for sub-genotype A1 were confirmed in the S and polymerase genes. The large number of complete genome sequences from sub-genotype A1 allowed for identification of additional amino acids and nucleotide residues characteristic of subgenotype A1. In particular nucleotide changes at position 1809/1811/1812 that alter the Kozak initiation sequence upstream the precore AUG, and A1888 in the core and precore ORF that creates an out of frame overlapping upstream ORF with the core ORF were found exclusively in subgenotype A1 isolates. Unique sequence changes in transcriptional regulatory elements were also identified in sub-genotype A1 isolates. The greater nucleotide divergence between subgenotype A1 isolates when compared with that of

subgenotype A2 isolates indicates that sub-genotype A1 has been endemic in the South African black population for a longer time than A2.

2.1. MATERIALS AND METHODS

2.1.1. SUBJECTS AND SERUM SAMPLES

Serum samples were collected from 275 HBsAg positive southern African Blacks: 260 were asymptomatic carriers (ASCs) and 15 were acute hepatitis B patients. The ASCs were randomly selected, unrelated Black factory workers and labourers from the Gauteng province in South Africa. The acute hepatitis patients were admitted and treated in Johannesburg Academic Hospital in recent years. The samples were stored at –70°C until analyzed. Commercially available kits (Abbott Laboratories, Chicago, IL) were used to detect HBV markers in the serum.

2.1.2 DNA EXTRACTIONS, AMPLIFICATION, CLONING AND SEQUENCING.

2.1.2.1. DNA EXTRACTION

Total DNA (genomic, mitochondrial, and viral) was extracted from the serum samples using the QIAamp DNA Mini Kit (QIAGEN GmbH., Hilden, Germany) according to the protocol provided by the manufacturer.

2.1.2.2. DNA AMPLIFICATION

Sub-genomic PCR amplification was initially performed to determine if the serum was HBV DNA positive and also used to determine if the sequenced clones were representative of the dominant HBV viral population in the serum of the individual.

To test for the presence of HBV DNA, a modification of the nested PCR reaction described previously was used (Kramvis *et al*., 1997). The reaction mixture for the first

round PCR consisted of 2.5 µl of 10× NH_4Cl buffer, 2 µl of 50 mM $MgCl_2$ buffer, 0.1 µl of HotStar Taq DNA polymerase (QIAGEN GmbH., Hilden, Germany) 5 units/ µl, 1.25 µl each of 20 µM primer 1687+ and 2498−, 2 µl of 40 mM dNTP mix and 2.5 µl of DNA made up to a final volume of 25 µl with distilled water cycling profile Table 2.1.

For the second round the reaction mix was made up of 5 µl 10× KCl buffer containing 15 mM of $MgCl_2$, 40 mM dNTP mix, 0.2 µl of HotStar Taq DNA polymerase 5 units/µl, 2.5 µl each of 20 µM primers 1730(+) and 2043(−), and 5 µl of first round PCR product made up to a final volume of 50 µl with distilled water for cycling profile see table 2.1. In cases where sufficiently high viral DNA was obtained the complete genome of the virus was amplified using the method described by Gunther *et al* (Gunther *et al*., 1995). Either the Expand High Fidelity System (Roche Diagnostics, Mannheim, Germany) or the TaKaRa Ex Taq system (TAKARA Biotechnology (Dailian) CO.,LTD, Shiga, Japan) was used.

When using the Expand High Fidelity System, the reaction mix consisted of 4.5 µl of 10× Expand High Fidelity (HF) buffer with 15 mM $MgCl_2$, 0.4 µl of 25 mM dNTP mix, 2.5 µl of 20 µM primer P1 and P2, and 5 µl of extracted DNA made up to 45 µl with distilled water. The enzyme mix was made up of 0.5 µl of HF buffer, 0.75 µl of Expand HF enzyme (3.5 units/ µl), and 3.75 µl of distilled water. The reaction mix was preheated to 94°C for 2 minutes and then cooled to 58°C at which point 5 µl enzyme

mix was added. This was followed by 40 cycles cycling profile shown in table 2.1. Both HBV positive and negative controls were included, the latter consisting distilled water instead of DNA in PCR reaction mixture. The reaction was carried out in the Perkin Elmer GeneAmp PCR System 9600 (Norwalk, CT) or Robocycler Gradient 40 (Stratagene, La Jolla, CA). The cycling conditions and reaction mix were the same when using the TaKaRa Ex Taq system.

When viral DNA concentration was low the complete genome was amplified using a modification of the method described by Takahashi *et al* (Takahashi *et al*., 1998). This involved the amplification of two overlapping fragments of the HBV genome 2.2 kb and 1.3 kb in length, respectively. This PCR was designed so that the overlap occurred over the variable regions of the S and precore/X genes, which would allow us to conclude that the amplified DNA was from a single genome, when overlapping regions were identical.

The reaction mix for the amplification of the 2.2 kb fragment (fragment B) consisted of 2.25 µl of 10× Ex Taq buffer with 20mM $MgCl_2$, 2 µl of dNTP (2.5mM each), 1.25 µL each of primers 1689(+) and 685(−), 2.5 µl of DNA, made up to 22.5 µl with distilled water. The enzyme mix was made up of 1.875 µl of distilled water, 0.25 µl of 10× Ex Taq buffer, and 0.375 µl of TaKaRa Ex Taq polymerase. The 22.5 µl reaction mix was preheated to 94°C for 2 minutes and 2.25 µl TaKaRa Ex Taq enzyme mix was added at the first annealing step. This was followed by 40 cycles, cycling profile (table 2.1). For the amplification of the 1.3 kb fragment (fragment A) using primers 1800(−) and

455(+) the annealing temperature was changed to 62°C for 30 seconds (table 1). Both HBV positive and negative controls were included. The former was a DNA extract from a known positive serum sample and the latter distilled water.

The PCR products for direct sequencing (used to determine the dominant viral population in circulation in the patient) were obtained from single round PCR using the primers 1763(+) and 1966(–). The reaction mixture was the same as for the second round PCR reaction above. A positive control and negative control was included in all PCR assays.

To avoid cross-contamination and false positive results the precautions and procedures suggested by Kwok and Higuchi 1989 were strictly adhered to (Kwok & Higuchi, 1989). DNA extraction, PCR reaction preparation, DNA Addition to PCR reaction, and electrophoresis were performed in separate areas. They were also designated areas for first and second round PCR.

2.1.2.3. DETECTION OF AMPLIFIED PRODUCT

For detection of the amplified product a 10 µl aliquot of the PCR product was electrophoresed on a 1 or 2% agarose gel containing ethidium bromide and visualized using an ultraviolet (UV) trans-illuminator. Positive and negative controls were electrophoresed in parallel with every assay. A 100 bp or 1 kb molecular weight marker (Promega., Madison, U.S.A) was also ran on each occasion to estimate the size of the PCR product.

2.1.2.4. CLONING

To ensure that there was enough DNA for sequencing and that a single viral genome was sequenced, amplicons were cloned into the pCR-Script vector (Stratagene, La Jolla, CA). All the steps involved in cloning were carried out according to the manufacturer's protocol. The amplified PCR product was purified of all contaminants. The ends of the PCR product where then polished with *Pfu* to remove sticky ends (to allow for blunt end ligation). This was followed by ligation of the blunt ended PCR product into the pPCR-Script Amp SK (+) cloning vector with T4 DNA ligase. The ligated product was transformed into competent *E coli* cells and plated onto ampicillin enriched agar plates containing X-gal and IPTG (appendix) for colour selection of positive clones.

For amplification of plasmids with HBV DNA inserts, positive colonies (colonies containing insert) were inoculated into 5 ml of LB medium containing ampicillin (appendix) and incubated at 37°C overnight while shaking at 175-200 rpm. Plasmid DNA was then extracted using the QIAprep Miniprep kit protocol (QIAGEN GmbH., Hilden, Germany). Before DNA extraction glycerol stocks were prepared for all bacterial cultures by adding 500 µl of the culture to an equal volume of autoclaved glycerol and these were stored at −70°C.

To identify vectors with inserts plasmid DNA was digested with restriction enzyme *Pvu* II (Amersham Pharmacia Biotech, Buckinghamshire, England) to yield a 2.6 kb plasmid DNA fragment and another fragment consisting of inserted DNA + 448 bp of

plasmid DNA. The digestion mix was made up of 0.5 µl of *Pvu* II (15 u/µl), 2 µl of *Pvu* II buffer, 5 µl of plasmid DNA made up to 20µl with distilled water. Digestion was carried out at 37°C for 2 hours and the product visualized under UV light after electrophoresis in a 1% agarose gel containing ethidium bromide. Plasmids selected for sequencing were grown in larger volumes of LB and the DNA extracted using the QIAfilter Plasmid Maxi Kit (QIAGEN GmbH., Hilden, Germany).

2.1.2.5. SEQUENCING

Selected positive clones and PCR product obtained from sub-genomic amplification were subjected to sequencing with the BigDye Terminator v3.0 Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, USA). The primers used for sequencing are listed in table 2.2. All sequences were analyzed in both the forward and the reverse directions.

2.1.3. PHYLOGENETIC ANALYSIS

Complete HBV genomes sequences were compared with corresponding sequences of HBV from GenBank. Multiple sequence alignments were carried out using Dambe (Xia and Xie, 2001). The alignments were edited manually in GeneDoc (Nicholas and Nicholas, 1997) and fed into PHYLIP (Phylogeny inference package) version 3.5c (Felsestein, 1995). DNAML (maximum likelihood) alone and DNADIST or PROTDIST consecutively with NEIGHBOR (neighbor-joining) were used to generate dendrograms. SEQBOOT, DNADIST and NEIGHBOR were used for bootstrapping of

1 000 data sets. CONSENSE was used to compute a consensus tree. Trees were visualized using TreeView Win 32 (Roderic D. M. Page, 2001) software programs.

Table 2.1. PCR Primers and Conditions Used for Amplification[a]

Primer	Amplification type	Position[b]	Sequence	Denaturation	Annealing	Extension	Size[c]	Reference
P1 P2	Complete genome	1821-1843 1825-1801	5'TTTTTCACCTCTGCCTAATCA3' 5'AAAAAGTTGCATGRTGMTGG3'	94°C 40 sec	58°C 90 sec	68°C 180 sec	3200	Gunther *et al*., 1995
455+ 1800–	Overlapping PCR	455-474 1800-1773	5'CAAGGTATGTTGCCCGTTTG3' 5'AGACCAATTTATGCCTACAGCCTCCTA3'	94°C 30 sec	62°C 30 sec	72°C 90 sec	1300	Takahashi *et al*., 1998
685- 1687+	Overlapping PCR	704-685 1687-1708	5'CGAACCACTGAACAAATGGC3' 5'CGACCGACCTTGAGGCATAC3'	94°C 30 sec	63°C 30 sec	2°C 120 sec	2198	Takahashi *et al*., 1998
1687+ 2498–	Sub- genomic	1687-1708 2498-2477	5'CGACCGACCTTGAGGCATAC3' 5'AAGCCCAGTAAAGTTTCCCACC3'	94°C 30 sec	62°C 40 sec	72°C 80 sec	812	Owiredu *et al*., 2001
1730+ 2043–	Sub- genomic	1730-1747 2043-2021	5'CTGGGAGGAGTTGGGGGA3' 5'CAATGCTCAGGAGACTCTAACGG3'	94°C 30 sec	62°C 50 sec	72°C 50 sec	314	Kramvis *et al*., 1997
1763+ 1966–	Sub-genomic	1763-1783 1946-1966	5'GGTCTTTGTACTAGGAGGCTG3' 5'GTCAGAAGGCAAAAACGAGAG3'	94°C 30 sec	58°C 50 sec	72°C 50 sec	204	Kramvis *et al*., 1997

(+), sense; (–), anti-sense.

[a] All amplifications began with an initial denaturation at 94°C for 2 minutes and ended with a final extension at 72°C for 10 minutes;

[b]Represents nucleotide position of HBV *ayw* (GenBank accession number AY233276) where the *EcoR*1 cleavage site is position 1

[c]Size of amplicon in base pair [genotype A]

[d]Extension time was increased by 120 seconds for every 10.

Table 2.2. Primers used for sequencing

Primer	**Sequence** 5' to 3'	**Position**[a]	**Annealing Temperature**	**Reference**
T3	5'TAATACGACTCACTATAGGG3'	772-791 on the PCR-Script vector	50°C	PCR-Script Amp cloning kit instruction manual.
T7	5'ATTAACCCTCACTAAAGGGA3'	627-646 on the PCR-Script vector	50°C	PCR-Script Amp cloning kit instruction manual.
2044+	5'CTCACCTCACCATACAGCACTCA3'	2044-2066	50°C	
1573-	5'CCGGCAGATGAGAAGGCACAGACGG3'	1573-1549	60°C	Owiredu *et al*., 2001
T733-	5'CCTGAGCCTGAGGGCTCCAC3'	3075-3094	55°C	Takahashi *et al*., 1998
2809+	5'CATCATTTTGTGGGTCACCAT3'	2809-2829	50°C	Owiredu *et al*., 2001
A100-	5'AGAGAAGTCCACCACGAGTCTAGA3'	270-247	55°C	Owiredu *et al*., 2001
192+	5'TCGTGTTACAGGCGGGGTTT3'	192-211	55°C	Takahashi *et al*., 1998
668-	5'GGCACTAGTAAACTGAGCCA3'	668-687	55°C	Takahashi *et al*., 1998
455+	5'CAAGGTATGTTGCCCGTTTG3'	455-474	55°C	Takahashi *et al*., 1998
455-	5'CAAACGGGCAACATACCTTG3'	475-455	55°C	
1156+	5'CGGCAACGGTCAGGTCTGT3'	1156-1174	55°C	Owiredu *et al*., 2001
2498-	5'AAGCCCAGTAAAGTTTCCCACC3'	2498-2477	55°C	Owiredu *et al*., 2001
A102-	5'GGTTGCGTCAGCAAACACTTGGCA3'	1197-1174	55°C	Owiredu *et al*., 2001
684+	5'TGCCATTTGTTCAGTGGTTCGTAGGGC3'	684-710	60°C	Owiredu *et al*., 2001
iS104	5'GCTCGGTACCAACTACTGTTGTTAGACGA3'	2335-2353	55°C	Owiredu *et al*., 2001
1763+	5'GGTCTTTGTACTAGGAGGCTG3'	1763-1783	58°C	Owiredu *et al*., 2001
1966-	5'GTCAGAAGGCAAAAACGAGAG3'	1966-1946	58°C	Owiredu *et al*., 2001

Abbreviations (+), sense; (−), anti-sense.

[a] Represents nucleotide position of HBV *adw* (GenBank accession number AY233276) where the *EcoR1* cleavage site is position 1 .

2.2. RESULTS

2.2.1. AMPLIFICATION, CLONING AND SEQUENCING

Two hundred and twenty three of the 260 DNA samples extracted from asymptomatic carriers and the 15 DNA samples extracted from acute HBV infected patients were positive for HBV DNA. Full-length amplicons were obtained from 24 of the 223 HBV DNA-positive asymptomatic carriers (twenty using the single round method described by Gunther *et al* and four using a modification of the method described by Takahashi *et al*). Five complete genomes were amplified from the fifteen HBV DNA positive samples from acute HBV patients using a modification of the Takahashi method. Of these eighteen full genome amplicons from ASCs and five from AH patients were successfully cloned and sequenced. These sequences have been deposited into GeneBank/EMBL/DDBJ database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed) as AY2333274 - AY233296. Direct sequencing of subgenomic PCR fragments confirmed that the clones represented the major viral strain in the serum of the individuals.

2.2.2. PHYLOGENETIC ANALYSIS

The length of the complete genomes of the seventeen genotype A isolates obtained was 3221bp. The insert of six nucleotides characteristic of genotype A was present in the core region of all seventeen isolates. The subtype of all seventeen isolates was deduced from the sequence to be *adw2* except for AY233288, which belonged to *ayw2*. The complete genomes and the four ORFs of these genotype A isolates were aligned with 31 sequences from GeneBank and phylogenetic analysis was carried out (figure 2.1 to

2.5). All isolates from acute HBV infected patients belonged to subgenotype A1, 10 isolates from asymptomatic HBV carriers belonged to subgenotype A1 and 2 to subgenotype A2. Division into subgenotype A1 and subgenotype A2 was evident in all 4 ORFs. Isolates from acute HBV infected patients could not be separated phylogenetically from asymptomatic carriers.

The intergroup divergence between subgenotype A1 and subgenotype A2 was 4.74% validating the existence of two different subgenotypes of genotype A (table 2.3). The intragroup divergence of subgenotype A1 isolates was greater than that of subgenotype A2 isolates (2.53% vs 1.39%) (table 2.3).

Table 2.3. Mean nucleotide divergence (%) of complete genome and individual open reading frame (ORF) sequences of HBV genotype A obtained using DAMBE.

	Intragroup			Intergroup
	Subgenotype A1	Subgenotype A2	Genotype A	A1vs A2
Complete Genome	2.53±1.00 (0.63-4.69)	1.39±0.89 (0.48-3.49)	3.36±1.47 (0.48-3.49)	4.74±0.56 (3.50-6.72)
Polymerase	2.52±0.95 (0.61-4.44)	1.23±0.70 (0.41-2.80)	3.35 ±1.48 (0.41-6.42)	4.77 ±0.49 (3.63-6.42)
X Gene	1.78±1.31 (0.00-6.45)	1.55 ±0.71 (0.00-3.01)	2.38±1.26 (0.57-6.45)	3.23 ±0.69 (1.72-4.95)
Precore/Core	2.40 ±1.19 (0.31-5.75)	1.75±1.67 (0.00-5.91)	3.26 ±1.65 (0.00-8.24)	4.58 ±1.08 (2.64-8.24)
Pre-S1/2	3.39 ±2.07 (0.43-9.07)	1.45±1.02 (0.43-7.68)	4.72±2.36(0.43-9.07)	6.99±1.02 (0.41-9.07)
HBsAg	1.21±0.56 (0.29-3.67)	0.84 ±0.62 (0.15-2.35)	1.53 ±0.74(0.15-3.67)	2.06±0.59 (1.03-3.67)

http://web.hku.hk/-xxia/software/software.htm. The sequences compared are those included in phylogenetic analysis (twenty one subgenotype A1 sequences and nine subgenotype A2 sequences).

¶ The mean nucleotide divergence (%)± and the range in parenthesis.

Figure 2.1-2.5. Dendrograms of complete HBV genomes and the four ORFs of genotype A isolates and representatives of the remaining seven genotypes. Subgenotype genotype A1 separates from subgenotype A2 in all ORFs although this is not strongly supported in the case of the precore and X ORFs (bootstrap value 54). Samples in red are South African isolates from the present study.

AS=isolates from asymptomatic carriers, AC= isolates from acute hepatitis patients, AF= isolates from fulminant hepatitis patients.

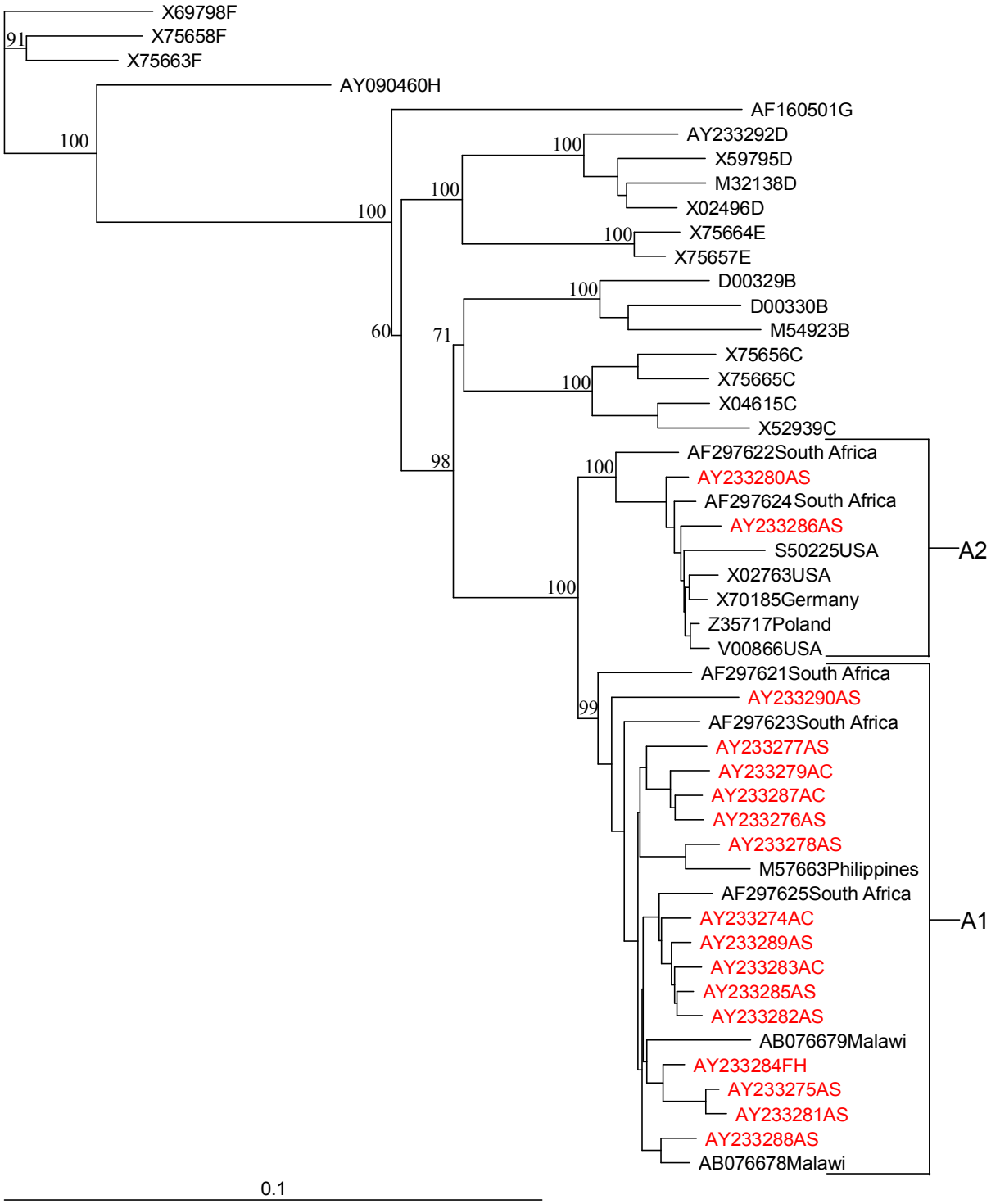


Figure 2.1. Complete genome

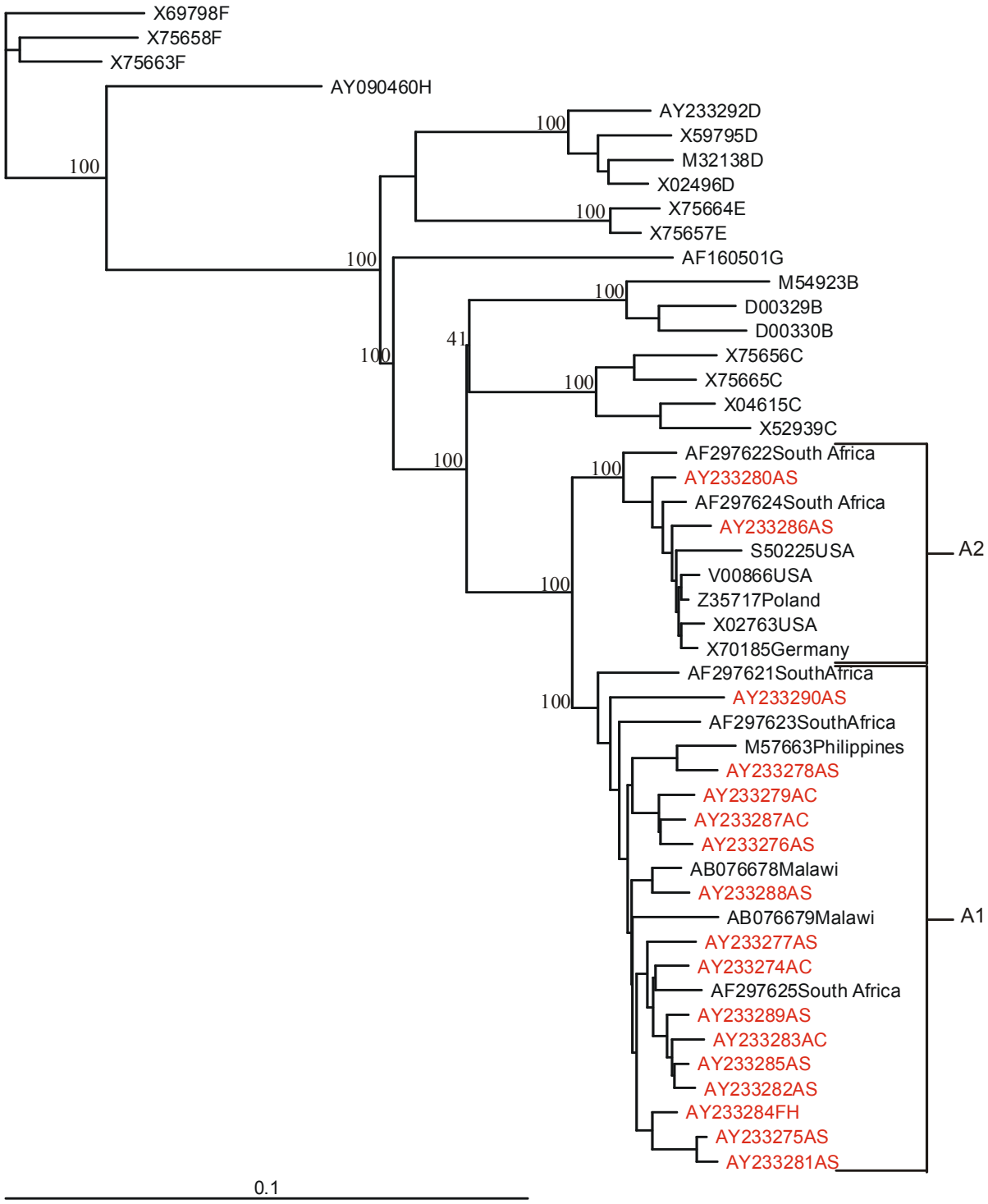


Figure 2.2. Polymerase gene

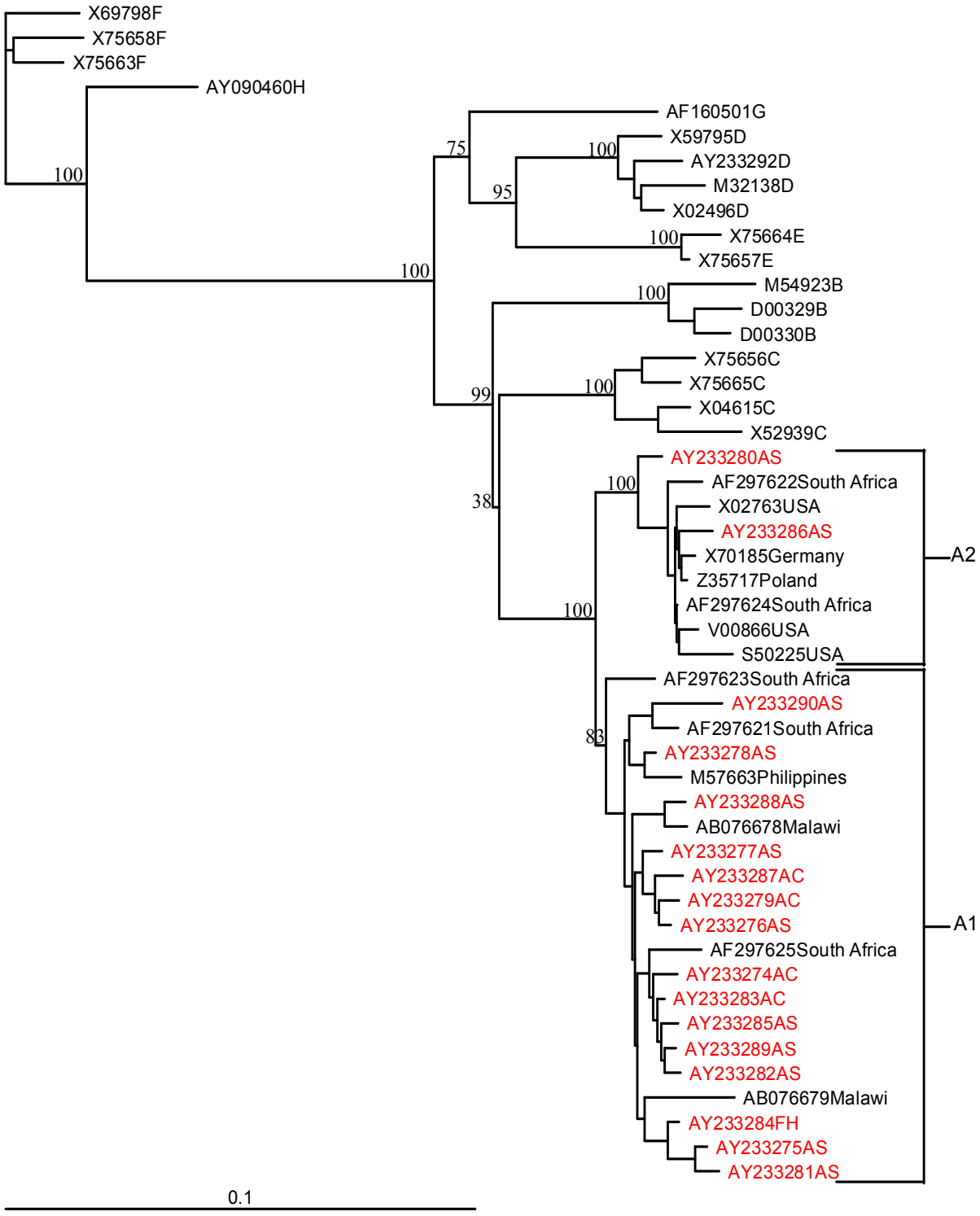


Figure 2.3. Complete S gene

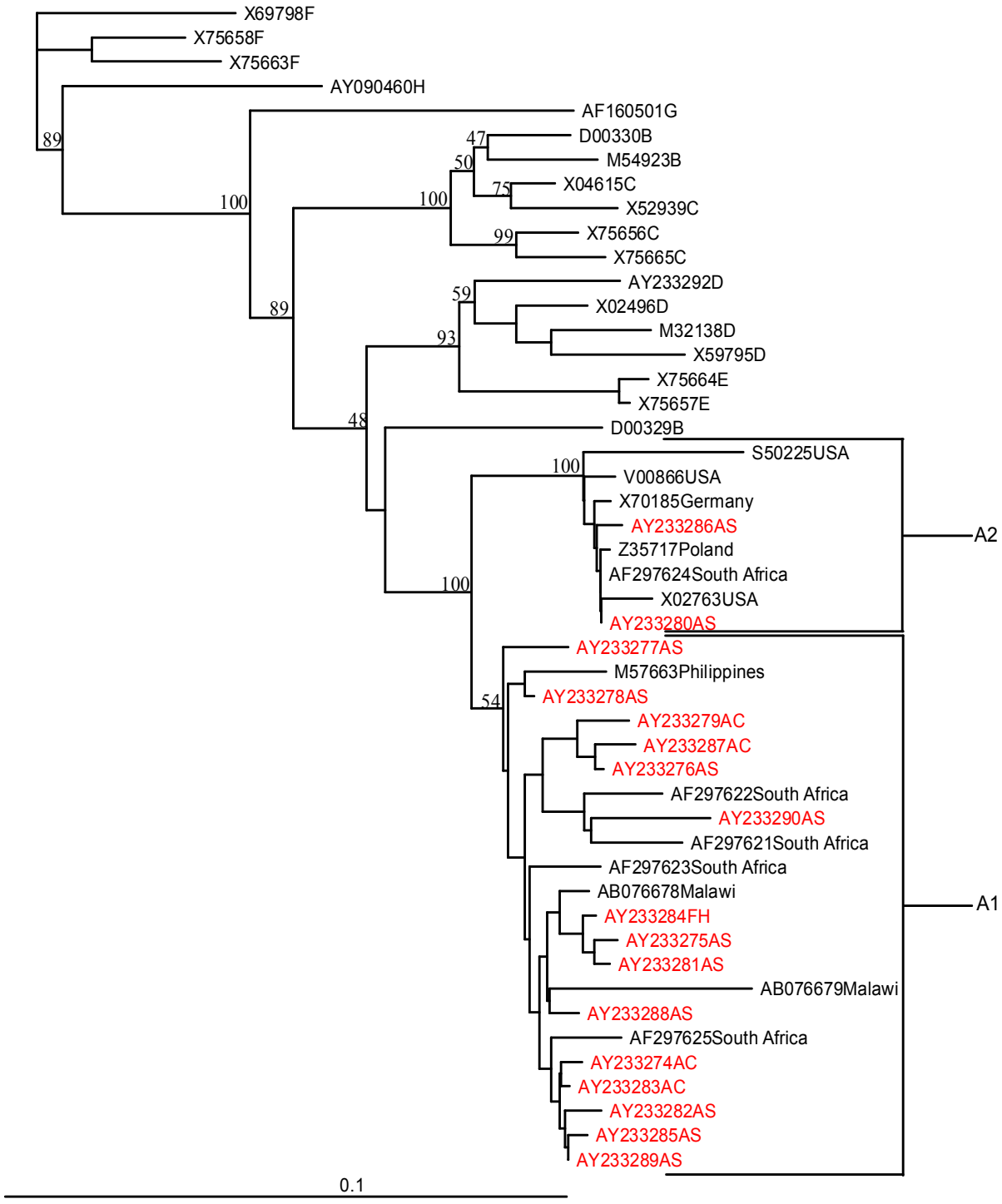


Figure 2.4. Precore/core gene

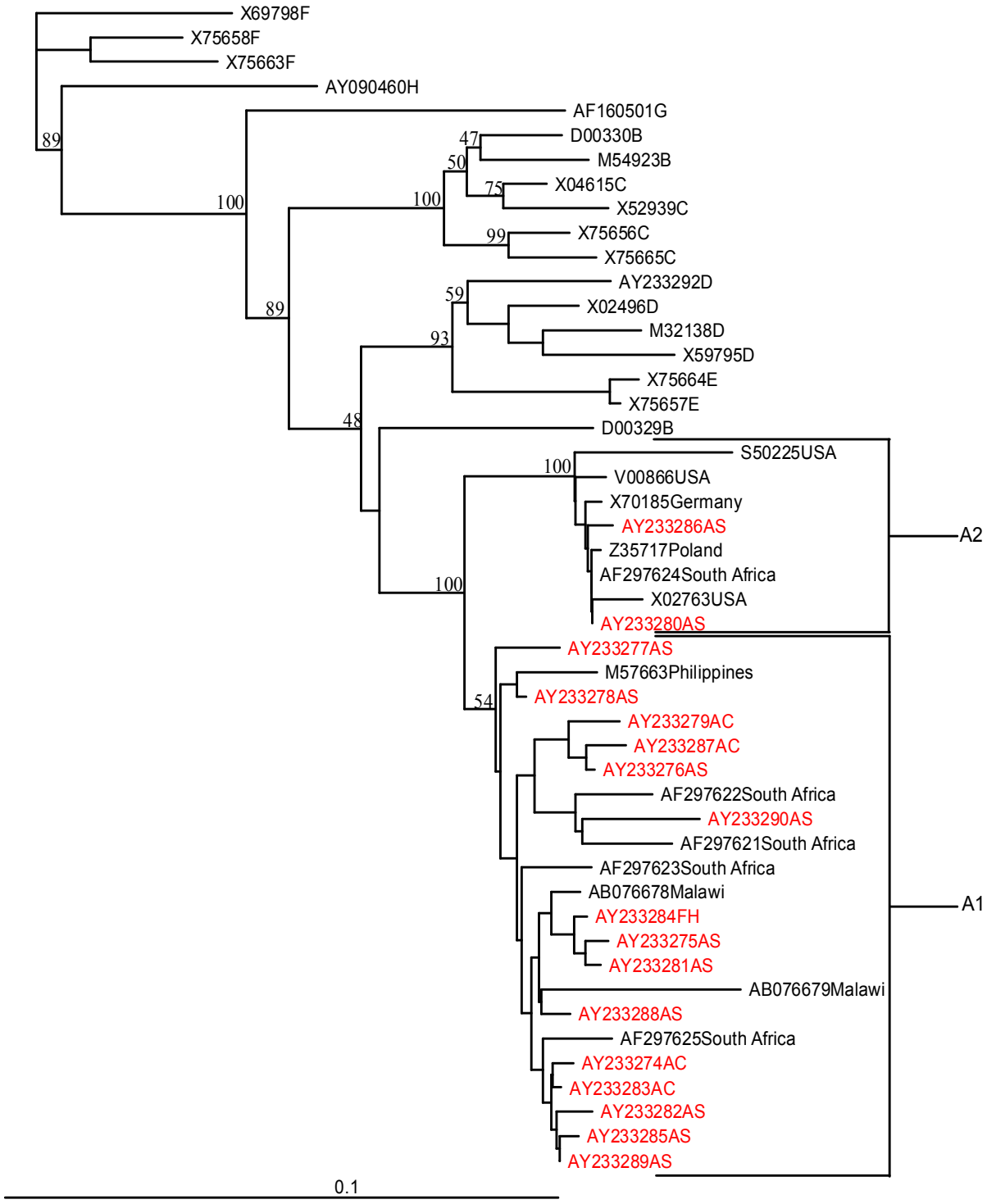


Figure 2.5. X gene

2.2.3. COMPARISON OF AMINO ACID SEQUENCES OF SUBGENOTYPE A1 WITH THOSE OF SUBGENOTYPE A2 AND OTHER GENOTYPES

Figure 2.6 provides a comparison of the translated sequences of the fifteen subgenotype A1 and two subgenotype A2 isolates to sequences obtained from GeneBank. The amino acids unique to subgenotype A1 are shaded in grey, those that are found in subgenotype A1 and other non-A genotypes, but not in subgenotype A2 are shown in bold. Amino acids Gln^{54}, Val^{74}, Ala^{86}, and Val^{91} in the pre-S1 region; Leu^{32} in the pre-S2; Thr^{236}, Gly^{268}, Tyr^{269}, $Gln^{334}$and Lys^{338} in the spacer of the polymerase gene; and Ser^{11}, Ala^{31}, Ser^{47}, Ser^{146}, Ser^{147} in the X ORF were unique to subgenotype A1. His^{182} and Ser^{251} in the polymerase ORF are shared by subgenotype A1 and genotype E. Leu^{54} in the pre-S2 is shared only with genotype D. By corollary, subgenotype A2 also had unique amino acids sequences that distinguish it from subgenotype A1 and all other genotypes. These are circled in figure 2.6.

Figure 2.6. Comparison of amino acid residues of S, polymerase and X ORFs of subgenotype A1 isolates with amino acids residues of subgenotype A2 and other HBV genotypes. Dots indicate amino acid identity and ? an amino acid deletion. Numbering according to *adw2* genome (GeneBank accession #AY233276). South African isolates sequenced in the present study are shaded in grey and all other sequences were obtained from GeneBank. For non-genotype A isolates the representative amino acid was obtained 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 genotype A1 and in other non-A genotypes but not in A2 are shown in bold, those found predominantly in subgenotype A1 and not in subgenotype A2 or in other genotypes are in bold and are shaded in grey. Amino acids unique to subgenotype A2 are circled.

*** Nucleotide positions according to the standardized numbering of the polymerase proposed by stuyver *et al* (Stuyver *et al*., 2001) are indicated in row three of figure 2.7.**

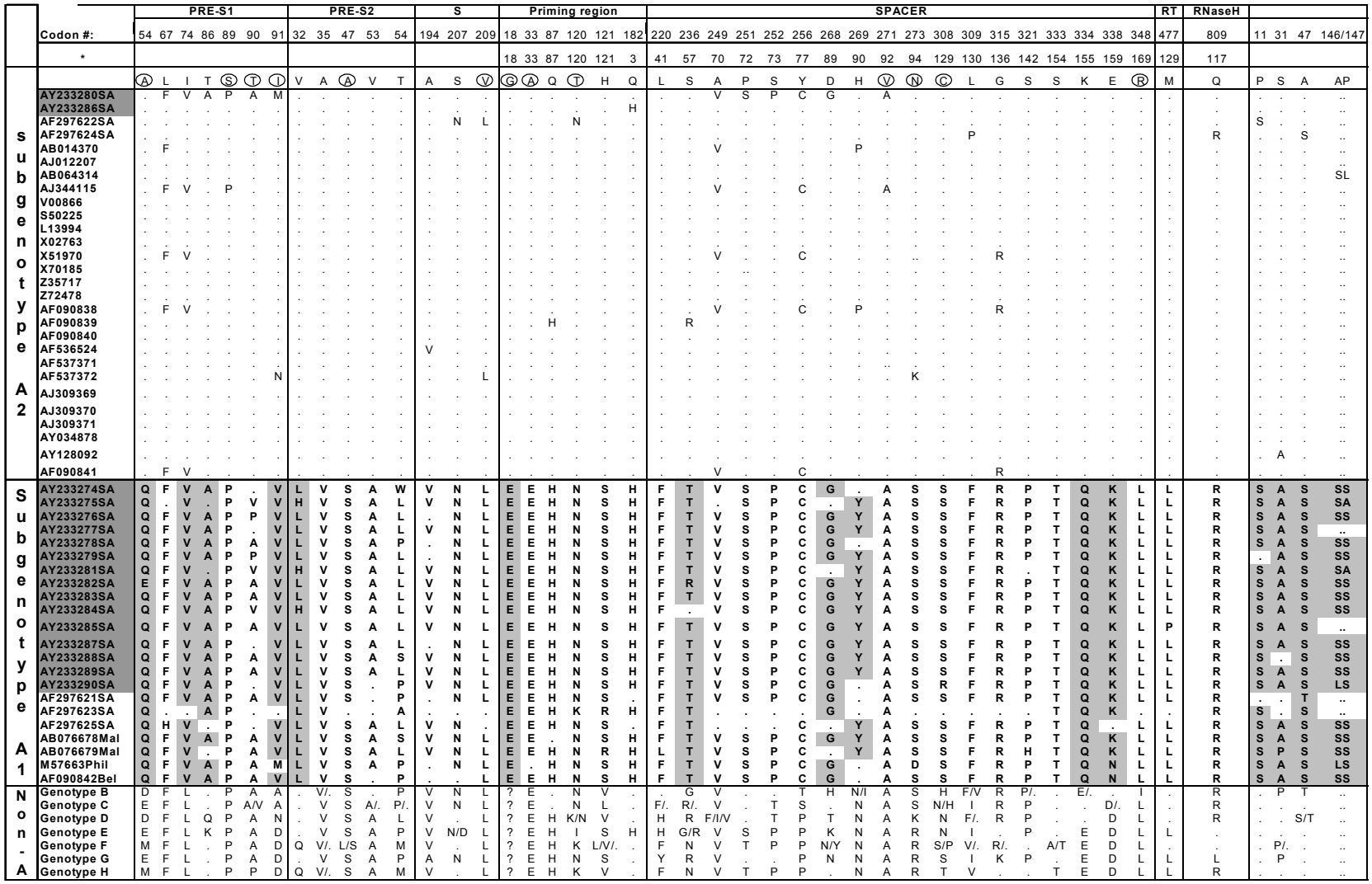

PRE-S1
PRE-S2
S
Priming region
SPACER
RT
RNaseH
Codon #:
Subgenotype A2
Subgenotype A1
Non-A
AY233280SA
AY233286SA
AF297622SA
AF297624SA
AB014370
AJ012207
AB064314
AJ344115
V00866
S50225
L13994
X02763
X51970
X70185
Z35717
Z72478
AF090838
AF090839
AF090840
AF536524
AF537371
AF537372
AJ309369
AJ309370
AJ309371
AY034878
AY128092
AF090841
AY233274SA
AY233275SA
AY233276SA
AY233277SA
AY233278SA
AY233279SA
AY233281SA
AY233282SA
AY233283SA
AY233284SA
AY233285SA
AY233287SA
AY233288SA
AY233289SA
AY233290SA
AF297621SA
AF297623SA
AF297625SA
AB076678Mal
AB076679Mal
M57663Phil
AF090842Bel
Genotype B
Genotype C
Genotype D
Genotype E
Genotype F
Genotype G
Genotype H

Figure 2.6

Figure 2.7. Comparison of the nucleic acid sequences of the cis-acting elements of subgenotype A1 isolates with sequences of subgenotype A2 and other HBV genotypes. Dots indicate amino acid identity. Nucleotide 1 denotes the nucleotide position of HBV *adw2* genome (GeneBank accession # AY233276) where the *Eco*RI cleavage site is at position 1. South African isolates sequenced in the present study are shaded in grey and all other sequences were obtained from GeneBank. For non-genotype A isolates the representative nucleotide sequence was obtained by aligning 10 sequences of each genotype from geographically distinct regions and the consensus nucleic acid deduced if it occurred in 60% or more of the sequences. Nucleotides found only in genotype A1 and in other non-A genotypes but not in A2 are shown in bold, those found predominantly in subgenotype A1 and not in subgenotype A2 or in other genotypes are in bold and are shaded in grey.

Figure 2.7

2.2.4. COMPARISON OF THE NUCLEOTIDE SEQUENCES OF *CIS*-ACTING ELEMENTS OF SUBGENOTYPE A1 WITH THOSE OF SUBGENOTYPE A2 AND OTHER GENOTYPES

The sequences of the *cis*-acting elements of the fifteen subgenotype A1 and two subgenotype A2 isolates were compared with sequences obtained from GeneBank (figure 2.7). Nucleotides unique to subgenotype A1 are shaded in grey, those found in subgenotype A1 and other non-A genotypes but not in A2 are shown in bold. Table 2.4 summarizes the mutations in the *cis*-acting regulatory elements characteristic of subgenotype A1 and the functional elements that are affected by these mutations.

Table 2.4. Mutations within the *cis*-regulatory elements found predominantly in subgenotype A1.

cis-acting element	Mutation	Functional element affected
Enhancer 1/X	A/T963C	5'modulator element
	C1092T	IRE
	T1155C	RFX1
Core promoter	C1404T	NRE-γ
	C/T1464G	NRE-β
	A/G1512T	NRE-β
	G1809T, A1811T, C1812T/G	Precore Kozak Sequence*
ε	G1888A	encapsidation signal
S1 promoter	T/G2720A	HNF-1 binding site
	[A/G3013, C/T3014A]	NF-1 transcription factor-binding site
S2 promoter	T/A/G3045C	region A
	C/A3109G, A/C3111T	region E
	[T3132C, C3133A]	region F

*Shown to reduce HBeAg translation by a ribosomal leaky scanning mechanism (Ahn *et al*., 2003). Functional domain according to Moolla *et al*., 2002, Nakao *et al*., 1999, and Garcia *et al*., 1993. Functional domains according to Moolla *et al*., 2002, Nakao *et al*., 1999, and Garcia *et al*., 1993.

2.3. DISCUSSION

There were no differences between isolates from asymptomatic carriers and those from acutely HBV infected patients. This finding suggests that differences in disease presentation may be the result of host factors, which may include immune state of the host at the time of infection, and host genetic factors. The following suggest that the host may play a very important rule in the outcome of HBV infection (Pontesilli *et al*., 2004, Toukan *et al*., 1990).

1. Some subjects in high risk groups i.e spouses in HBV infected family do not get infected, indicating the presence of some sort of genetic defense mechanism against the virus.
2. Prevalence of chronic and acute HBV infection differs between populations from different geographical locations. This however may be due to genotype differences.
3. Response to antiviral therapy differs in individuals even when they are infected with identical strains.
4. Vaccination is not completely protective in a proportion of those vaccinated.

The characterization of fifteen subgenotype A1 and two subgenotype A2 isolates from asymptomatic carriers and acutely HBV infected patients allows for confirmation of previous findings regarding subgenotype A1 (previously called subgroup A') (Bowyer *et al*., 1997, Kramvis *et al*., 2002) as well as a more detailed analysis of subgenotype A1 from Africa and identification of features unique to this subgenotype.

All genotypes A isolates sequenced in this study had a genome length of 3221 and did not have the 21bp deletion, that was found previously in isolates from fulminant hepatitis patients from the same region (Oweridu *et al*., 2001b). This deletion was also absent from subgenotype A1 isolates from chronic carriers from Malawi, Philippines, India, Nepal, and Bangladesh (Sugauchi *et al*., 2004a, Sugauchi *et al*., 2003). A clear separation of subgenotype A1 and A2 can be seen at the level of the complete genome and in all four ORFs (figure 2.1-2.5).

The amino acids differentiating the subgenotypes of genotype A from each other and from other genotypes were concentrated in the pres-S1 region overlapping the spacer of the polymerase (figure 2.6). The sequence of the pre-S1 region has been shown to be well conserved within a given HBV subtype (Uy *et al*., 1992) and this region may play a role in the attachment of HBV to hepatocytes (Neurath *et al*., 1986, Pontisso *et al*., 1989). Therefore, it is possible that the molecular evolution of the pre-S1 sequence is determined by the host population (Kramvis *et al*., 2002).

$Gln^{334}$and Lys^{338}, which are unique to subgenotype A1, are found in the fingers of the HBV polymerase within the positively charged DNA-binding cleft (Das *et al*., 2001) (figure 2.8) Gln^{334} is uncharged and replaces basic Lys in subgenotype A2 and negative Glu in other genotypes. On the other hand, positively charged Lys^{338} replaces Glu and Asp, both of which are negatively charged and found in genotype A2 and the other genotypes respectively. The change in charge caused by the alternate amino acids in this region could possibly affect the binding of the DNA to polymerase and result in a

decrease in viral DNA synthesis. In fact, subgenotype A1 has been associated with low HBV DNA levels (Kramvis *et al*., 1997) and comparison of subgenotype A1 with A2 and genotype D has revealed that HBV carriers infected with subgenotype A1 have significantly lower levels of HBV DNA (Tanaka *et al*., 2004).

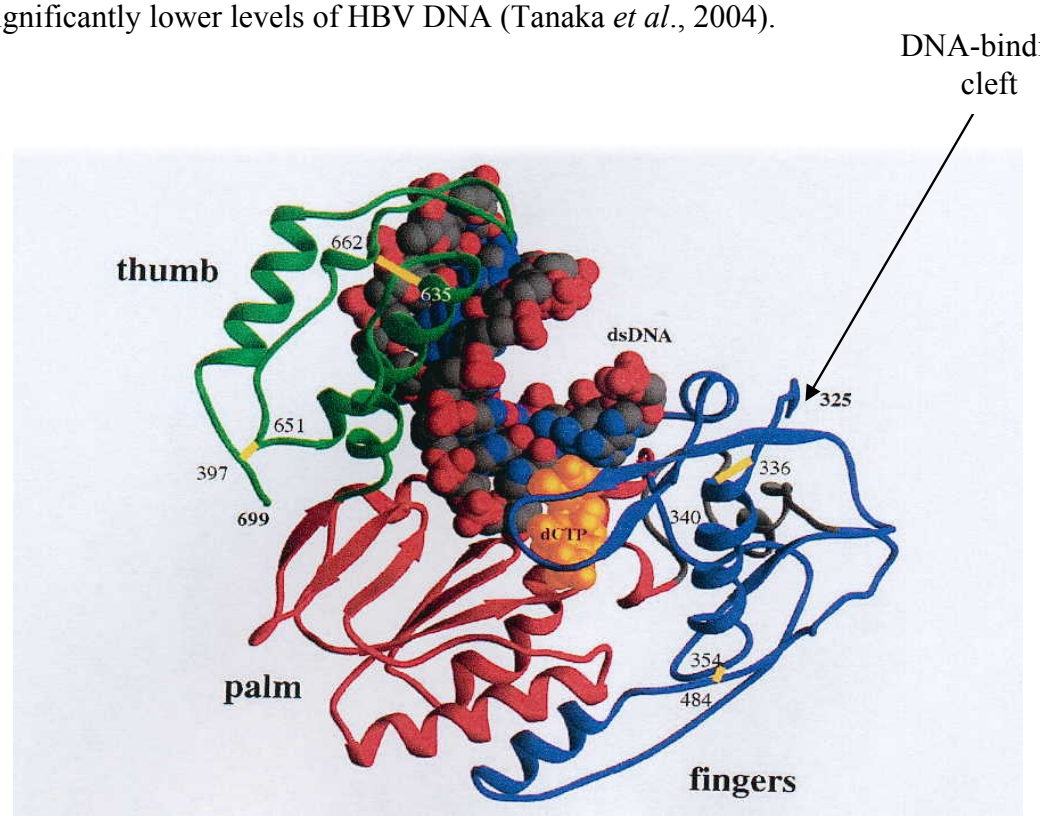


Figure 2.8. HBV polymerase represented as a ribbon diagram with fingers 325-403 and 469-519, palm 404-440 and 520-613, thumb 614-699 sub domain in blue, red and green respectively. The bound dsDNA template primer is shown as a space-filled in grey (with Nitrogen and Oxygen atoms in blue and red respectively. The four-disulphide links are represented in yellow (Das *et al*., 2001).

The following signature amino acid motif, Ser^{11}, Ala^{31}, Ser^{47}, Ser^{146} and Ser^{147}, was unique to subgenotype A1. These changes were found in regions of the HBx protein that are not functionally active in transactivation (Arii *et al*., 1992). Ser^{146}, Ser^{147} in the X ORF are a result of mutations at nucleotide position 1809 and 1812.

The double mutation at A1762T/G1764A was more common in subgenotype A2 isolates that A1 isolates: 14/28 vs 3/22 (figure 2.7). Two isolates from asymptomatic carriers in the present study had this mutation. The A1762T/G1764A mutations have been reported to occur at a significant higher rate in HCC patients when compared with asymptomatic carriers (Baptista *et al*., 1999). Hence it is not surprising that it was found only in ASCs from the present study. These mutations, because they occur within the binding site of the liver transcription factors reduce the amount of HBeAg produced by reducing the amount of precore mRNA (Buckwold *et al*., 1996, Gunther *et al*., 1998, Moriyama, 1997).

Double or triple point mutations at position 1809-1812 were found only in subgenotype A1 isolates (figure 2.7). In a previous study by Baptista *et al* it was reported that 80% of South African isolates harbor similar mutations immediately upstream of the precore AUG codon (Baptista *et al*., 1999), which might impair translation initiation of HBeAg (Kozak, 1986b, Kozak, 1987b, Kramvis and Kew, 1999). This hypothesis was tested and it was shown that HBeAg expression was severely impaired by 1809T 1811T 1812T (MT1) and 1809T 1811C 1812T (MT2) triple mutations and moderately reduced by 1809T 1812T (MT4) and 1809A 1812T double mutations (MT5) (Ahn *et al*., 2003).

Table 2.5. Different types of mutations at 1809/1811/1812 affecting HBeAg synthesis.

Mutation	Sequence	% Reduction in HBeAg synthesis
WT	**G**CA^{-3}**C**CATGC	0%
MT1	**T**C**T**$^{-3}$**T**CATGC	83%
MT2	**T**CA**G**CATGC	0%
MT3	**T**C**C**$^{-3}$**T**CATGC	76%
MT4	**T**CA**T**CATGC	20%
MT5	**A**CA**T**CATGC	22%

*WT represents the wild-type of subgenotype A2. MT1 and MT3 have the most effect because they involve a change at the -3 position (Ahn *et al*., 2003).

MT4 was the most common mutation in the subgenotype A1 isolates sequenced in the present study, being present in 10 of 15, followed by MT2, 2/15, and WT, each present in 2 of 15. MT1 was present in 1 of 15. The representation was similar when analysis was carried out on 55 subgenotype A1 precore sequences from GeneBank MT4 in 42 of 55, MT2 in 4 of 55, MT3 in 3 of 55, MT1 in 1 of 55, MT5 in 1 of 55 and WT in 4 of 55. It is also worth mentioning that of the 55 subgenotype A1 isolates analyzed, only 4 isolates from South Africa had a wild type (WT) at 1809/1812 (4/55). This was also the case in previous studies (Baptista *et al*., 1999, Kramvis *et al*., 1997, Kramvis *et al*., 1998).

Table 2.6. Representation of MT matutions in subgenotype A1 isolates.

Mutation	Relative propotion in 15 subgenotype A1 isolates from present study	Relative proportion in 55 subgenotype A1 isolates from GeneBank
WT	2/15 (13%)	4/55 (7%)
MT1	1/15 (6%)	1/55 (2%)
MT2	2/15 (13%)	4/55 (7%)
MT3		3/55 (5%)
MT4	10/15 (66%)	42/55 (76%)
MT5		1/55 (2%)

*MT4 is the most abundant mutation.

These mutations are not a result of adaptive change under immune pressure because they are found in HBV isolates obtained from children and in acutely infected HBV patients (Ahn *et al*., 2003).

A G to A mutation at nucleotide 1888 was unique to subgenotype A1 isolates (figure 2.7). In addition to stabilizing the encapsidation signal and possibly affecting reverse transcription this mutation may affect the translation of the core protein.

Regulatory *cis*-acting elements overlap with the protein coding ORFs of HBV. Therefore, changes in the protein-coding region can also lead to alterations in these elements. Table 2.4 shows the mutation within the regulatory elements that are

characteristic of subgenotype A1 and the regions of the *cis*-acting elements affected by these mutations. Although it is not possible to deduce the effect of these changes on the replication of the virus, it is probable that they have a role to play in modulating replication and resulting in the reduced HBV DNA levels that are observed in patients infected with subgenotype A1 (Kramvis *et al*., 1997, Tanaka *et al*., 2004).

2.4. CONCLUSION

They were no characteristic differences between isolates from asymptomatic carriers when compared with those from acutely HBV infected patients. The detailed analysis of the seventeen genotype A isolates from the present study with isolates from other regions of the world have confirmed many previously reported differences between subgenotype A1 and A2 and identified new ones. Subgenotype A1 isolates from South Africa differ from western subgenotype A2 isolates in two ways. Firstly, subgenotype A1 isolates have distinctive sequence characteristics that may affect both viral replication and protein expression. Secondly, the mean nucleotide divergence of subgenotype A1 is greater than that of subgenotype A2, suggesting that subgenotype A1 has been endemic and has a long natural history within the South African black population.

New studies have identified subgenotype A1 isolates in the Philippines, India, Nepal, and Bangladesh (Sugauchi *et al*., 2004a), Malawi (Sugauchi *et al*., 2003) and in communities with continuing African links in Yemen (Sallam and William Tong, 2004). This wide geographic distribution together with a higher intragroup divergence of subgenotype A1 as compared to subgenotype A2 isolates may indicate that subgenotype A1 is phylogenetically much older (Sugauchi *et al*., 2004a).

CHAPTER 3

EFFECT OF THE G1888A MUTATION OF SUBGENOTYPE A1 ON THE TRANSLATION OF THE CORE PROTEIN

SUMMARY

A distinctive characteristic of subgenotype A1 isolates of hepatitis B virus is the presence of a G-to-A substitution at nucleotide 1888 of the precore/core gene. This nucleotide substitution introduces an out-of-frame AUG, creating an overlapping uORF, which terminates five nucleotides downstream from the core AUG. This uORF can potentially be translated into a seven amino acid peptide: 'Met-Ala-Leu-Gly-His-Gly-His'. In addition to stabilizing the encapsidation signal by forming a Watson and Crick base pair with T1871, this mutation may affect translation of the core protein, most probably by a leaky scanning mechanism. The aim of this study was to determine whether this mutation had any modulating effect on the translation of the core protein, and if, so the mechanism involved. The complete core gene together with part of the precore of sub-genotype A1 was cloned into the amino terminal of a green fluorescent protein (GFP) plasmid. This enabled quantification of the expression of the core/GFP fusion protein by flow cytometry. The introduction of the uORF resulted in an 18.75% reduction of core gene expression. This reduction was more pronounced (64.84%) when the suboptimal Kozak sequence of the 1888 AUG was replaced with an optimal Kozak sequence. This further reduction in the expression of core protein intimates that the 18.75% reduction is as a result of leaky scanning. By increasing the distance between the stop of the overlapping uORF and the core AUG by a minimum of 15 nucleotides core/GFP expression was almost double on average, indicating that stalling

of ribosomes at the stop of the uORF may be interfering with initiation at the core AUG through a steric hindrance mechanism. Our findings indicate that the G1888A mutation decreases core protein production by a leaky scanning mechanism. This decrease may be important in minimizing the immune response during initial infection, predisposing the individual to a chronic state of infection.

3.1. INTRODUCTION

The HBV core protein, which forms the building block of the core particle, is translated from the pregenomic RNA (Tiollais *et al*., 1985). A stem loop secondary structure at the 5' end of the pregenomic RNA known as the encapsidation signal (ε) is required for encapsidation of the viral pregenome prior to viral replication. Binding of viral polymerase to ε is required for initiation of the encapsidation reaction as well as DNA priming (Nassal and Schaller, 1993). The encapsidation signal also functions in activation of the viral polymerase and template restriction during viral replication (Kramvis and Kew, 1998). Some amount of the core protein can be produced from the precore mRNA transcript through a leaky scanning mechanism (Ahn *et al*., 2003). The core protein together with the viral polymerase and viral pregenome comprise the core particle, which is the location for HBV replication.

Data in the previous chapter identified a unique G to A mutation at position 1888 in subgenotype A1 isolates (G1888A). In addition to stabilizing the encapsidation signal by forming a Watson and Crick base pair with T1871, this nucleotide substitution introduces an out-of-frame AUG creating an overlapping uORF, thirteen nucleotides from the core start AUG, and terminating five nucleotides downstream from the core start codon.

Upstream AUGs can affect translation or function of downstream proteins by one of the following mechanism;

1. Leaky scanning: leaky scanning occurs when uAUGs are not recognised 100% by the scanning initiation complexes, hence some of these will bypass the uAUGs and initiate at the next downstream AUG (Kozak, 1978).
2. Re-initiation: this occurs when after translation of the uORF, 40s ribosomal subunits remain attached to the mRNA and resume scanning at the start of the next downstream ORF (Kozak, 1999).
3. Production of an N-terminally extended protein: this occurs when the introduced uORF is in-frame with the downstream major ORF (Calkhoven *et al*., 2000).

The uORF created by the newly introduced AUG at 1888 is out of frame with the core AUG. This eliminates the possibility that re-initiation, or production of an N-terminally extended protein may affect translation or function of the core protein, respectively.

The context surrounding the AUG initiation codon plays a vital role in its recognition. The optimal recognition context for vertebrates is GCCR^{-3}CCAUGG^{+4}, the purine at -3 and the G at +4 are the most important contributors with an A preferred to a G at -3 (Kozak, 1986a, Kozak, 1997). A strong sequence context contains both the R at -3 and a G at +4, whereas a satisfactory sequence context has only one of them. Greater than 90% of AUGs initiating major ORFs have either a satisfactory or strong consensus as compared to less than 50% of those initiating uORFs (Kozak, 1987b, Suzuki *et al*., 2000). The context of the core AUG is strong, with a G at -3 and a G at +4, while that

of the 1888 AUG is satisfactory with a T at -3 and G at +4. AUG recognition may also be affected by the position of the AUG i.e if the AUG occurs within a secondary structure it's recognition might be impaired, while recognition on a linear stretch of mRNA is optimal (Kozak, 1986a).

In the present study it was investigated whether the G1888A mutation had any effect on the translation of the downstream core protein using a core/ green fluorescent protein fusion protein (core/GFP). The core/GFP fusion protein allowed for detection of minor differences, which could otherwise not be detected by other molecular techniques i.e *in vitro* coupled transcription/translation, and western blotting.

3.2. MATERIALS AND METHODS

3.2.1. AMPLIFICATION OF THE PRECORE AND CORE GENE

The complete precore gene and core gene of subgenotype A1 was amplified using the following conditions. The PCR reaction mix consisted of 5 µl of Expand High Fidelity buffer 2, 4 µl of dNTP mix (2.5mM each), 0.5 µl of Expand High Fidelity enzyme (3.5units/µl), 2.5 µl each of 20µM forward and reverse primer, 1 µl of appropriately diluted DNA, made up to a final volume of 50µl with best quality water (BQW). The PCR reaction consisted of initial denaturation for 2 min at 95ºC, followed by 40 cycles 94ºC for 30secs, 62ºC for 30sec, 72ºC for 1 min and a final extension of 72ºC for 10 min. The amplified product was then visualized on a 1% agarose gel stained with ethidium bromide.

3.2.2. *IN VITRO*-TRANSCRIPTION/TRANSLATION

The plasmids constructed for use in *in vitro*-transcription/translation studies are shown in table 3.1. The reaction conditions for the amplification of the precore or core gene are described in 3.2.1. The amplified PCR products were ligated into the pCR-Script Amp cloning vector using the Rapid Ligation Kit (Roche). The ligated product was transformed into competent *E coli* cells and plated onto ampicillin enriched agar plates. These plates where incubated at 37°C overnight, after which colonies where selected and inoculated in ampicillin enriched LB for plasmid DNA amplification. Positive clones were screened by restriction enzyme digestion with *Xho*1 and *Eco*R1 and confirmed by sequencing.

Table 3.1. Plasmid used for *in vitro* translation.

Plasmid	1809/1812	1888	Template	Primer
precore1	**T**CA**T**	**A**	Subgenotype A1	MT1and Rev
precore2	**T**CA**T**	**G**	Mutagenesis of precore 1	S1888F and S1888R
Precore control	**G**CA**C**	**G**	precore 2	WT and Rev
core1	N/A	**A**	precore 1	MTA and Rev
core2	N/A	**G**	precore 2	MTG and Rev

* Precore 1, precore 2, precore control contain the precore gene cloned into the *Xho*1 and *Eco*R1 site in pPcr-Script Amp cloning vector while core 1 and core 2 contained complete core gene cloned in the *Xho*1 and *Eco*R1 of the vector.

Coupled transcription/translation was performed at 30ºC using the TNT Quick Coupled Transcription/Translation Systems (Promega), with [^{35}S]-methionine. The T7 promoter on the pCR-Script vector was use to drive transcription. An aliquot of the translation product was electrophoresed in a 15% polyacrylamide gel followed by treatment with enlightening solution (Amersham Biosciences). The gel was then dried at 80ºC and exposed to X-ray film at -80ºC.

3.2.2. CONSTRUCTION OF GFP FUSION PLASMIDS

GFP fusion plasmids were obtained by the fusion of the complete core gene of subgenotype A1 to the amino terminal of the EGFP on the pWay20 plasmid (generously provided by Thomas Hughes, Molecular Motion Lab Montana State University). The PCR product was cloned into pWay20 using the Strategene pCR-

Script Amp cloning kit according to manufacturer's instructions to produce core A1. Positive clones were screened by restriction enzyme digestion with *Eco*R1 and confirmed by sequencing.

Table 3.2. Primers used for plasmid construction.

Primer	Sequence 5'-3'
S1888F	TTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTT
S1888R	AGCTTGGAGGCTTGAACAGTGGGACATGTACAA
CoreF	CACCTCGAGACTTTTTCACCTCTGCCTAATCATC
CoreR	AGTACCTCCACCTCCACTACCTCCACCTCCACATTGAGATTCCCGAGATTG
MT1	GGAGGCTCGAGGCATAAATTGGTCTGCGCACCATCATCATG
WT	GGAGGCTCGAGGCATAAATTGGTCTGCGCACCAGCACCATG
MTA	GGAGGCTCGAGGAAGCTGTGCCTTGGATGGCTTTGGGGCATG
WTG	GGAGGCTCGAGGAAGCTGTGCCTTGGGTGGCTTTGGGGCATG
Rev	AAAGCGAATTCAAGTTTCCCACCTTATGAGTCC
GFP1-	CACCATGGTGGCGACCCATCCAAGAAGCCGAATTCCAC
GFP1+	GGTCGCCACCATGGTGAGTAAGGGCGAGGA
1888kozF	GCCTCCAAGCTGTGCCACCATGGCTTTGG
1888KozR	GGCACAGCTTGGAGGCTTGAACAGTG
P15F	GCTTTGGGGCATGGACATAGACCCTTACAA
P15R	ATGTCCATGCCCCAAAGCCATGGTGGG
P60F	ACTCTCATTTTTGCCTTCAGACTTCTTTC
P60R	GAAGGCAAAAATGAGAGTAACTCCACA
P100F	GAAGCCTTAGAGTCTCCAGAGCATTGCT
P100R	GGAGACTCTAAGGCTTCCCGATACAGTG

***** The nucleotides in red (CoreR) code for the glycine chain introduced between the core gene and the GFP coding sequence of pWay 20.

3.2.3. SITE DIRECTED MUTAGENESIS

Site directed mutagenesis was used to create core/GFP fusion plasmids in which the stop codon of the uORF was extended (table 3.3.). Three fusion plasmids were produced: one with the uORF stop codon 15 nucleotides from the core AUG (G1888A (15)), another with the uORF stop codon 60 nucleotides from the core AUG (G1888A (60)), and the third with the uORF 100 nucleotides from the core AUG (G1888A (100)). The uORF stop codon in the plasmid with the G1888A mutation is 5 nucleotides from the core AUG (G1888A (5)).

CoreA1+Opt was constructed to introduce an optimal Kozak initiation sequence around the 1888 AUG. Core A2 was obtained by changing the G1888A mutation of core A1 to wild-type G1888. UpWay19 was obtained by introducing a uORF (similar to that introduced by the G1888A mutation) upstream of the wild-type GFP AUG.

Site directed mutagenesis was performed using two overlapping primers (www.invitrogen.com) with one or both of them containing a target mutation to generate various products (figure 3.1). The PCR reaction mix consisted of 5 µl of Expand High Fidelity buffer, 4 µl of dNTP mix (2.5mM each), .5 µl of expand high fidelity enzyme (3.5units/µl), 2.5 µl each of 20µM forward and reverse primer (table 3.3), 1 µl of appropriately diluted DNA, made up to a final volume of 50 µl BQW. The PCR reaction consisted of initial denaturation for 2 min at 95ºC, followed by 25ºcycles 94ºC for 30secs, 55ºC for 30sec, 68ºC for 6min (1 kb=1 min) and a final extension of

68ºC for 10 min. 10 µl amplified product was then visualized on a 1% agarose gel stained with ethidium bromide.

The remaining amplicon (40 µl) was then run on an agarose gel stained with 50 µl of 2mg/ml crystal violet. The crystal violet migrated in a direction opposite to that of DNA, following electrophoresis the gel was placed on a light box and the band of interest excised and purified with the S.N.A.P. UV-free Gel Purification Kit (Invitrogen Life Technologies). The purified PCR product was transformed into competent *E coli* cells and plated onto ampicillin enriched agar plates. These plates where incubated at 37°C overnight, after which colonies where selected and inoculated in ampicillin enriched LB for plasmid DNA amplification. Mutations where confirmed by sequencing.

Table 3.3. Primers sets used for site directed mutagenesis.

Primer set	Template	Product obtained
GFP+ and GFP-	pWay19	UpWay19
S1888F and S1888R	CoreA1	CoreA2
1888kozF and 1888KozR	G1888A	CoreA1+Opt (G1888A5)
P15F and P15R	G1888A5	G1888A15
P60F and P60R	G1888A15	G1888A60
P100F and P100R	G1888A60	G1888A100

*The template and the products obtained are also indicated

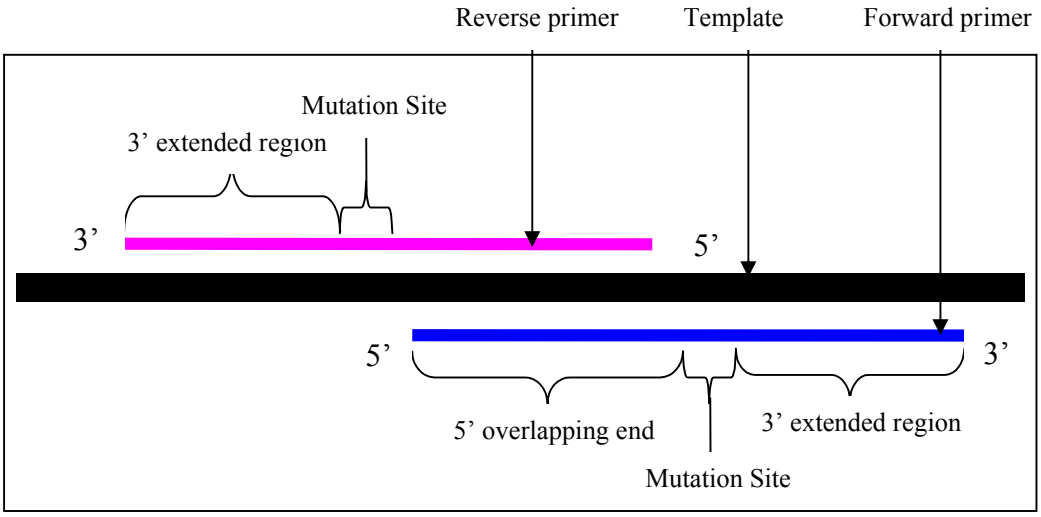


Figure 3.1. Diagrammatic representation of primer design for site directed mutagenesis www.invitrogen.com. The reverse primer is represented in pink and forward primer in blue. The primer pair is made up of a 15-20 bp overlapping region for optimal ligation of mutagenic plasmid ends. The mutation site can made up of up to a 21 bp deletion, substitution, and or insertions. The 3' extended region of the forward primer should be made up at least 10 bp from the mutation site for adequate annealing of the primer.

3.2.4. TISSUE CULTURE AND TRANSFECTION

The human hepatoma cell line HuH7 a generous gift from Dr H Nakabayahsi, Nagoya City University, Graduate school of medical sciences, Nagoya, Japan were maintained in RPMI 1640 + L-Glutamine medium (Invitrogen Life Technology). The medium was supplemented with 10% (v/v) foetal bovine serum. The cells were grown at 37ºC in a humidified atmosphere containing 5% CO_2. Cells were plated in 60 mm Petri dishes 24 hrs before transfection. Transfection was carried out at 50% confluency using

lipofectamine™ (Invitrogen Life Technology) as described by the manufacturer. 24hrs prior transfection 4 x 10^4 cells were seeded into 60mm petri dishes so that they were 50% confluent on the day of transfection. Cell growth medium was changed 4-6 hrs before transfection. Cells were plated in 5 mls of their normal medium supplemented with serum and without antibiotics. For each Petri dish of cells to be transfected 8µg of plasmid DNA was added to 500 µl of OPTI-MEM® (Invitrogen Life Technology). 4 µl of lipofectamine™ was diluted in 500 µl of OPTI-MEM® and incubated for 5 mins at room temperature. The diluted lipofectamine™ was then combined with the plasmid DNA and incubated for 20 mins, to allow for formation of DNA lipofectamine™ complexes. A 1000 µl DNA- lipofectamine™ complex was then added to each Petri dish, followed by a 48 hrs incubation at 37ºC in a humidified atmosphere containing 5% CO_2.

3.2.5. FLOW CYTOMETRY

3.2.5.1. PREPARATION OF CELLS FOR FLOW CYTOMETRY

48 hrs post-transfection, HuH7 cells transfected with Core/GFP fusion plasmids, were rinsed once with 3 ml PBS, followed by incubation at 37ºC for 10 min in 3 ml of 0.1% EDTA. The cells were then suspended and transferred into a 15 ml tube. The EDTA action was stopped by the addition of 300 µl of serum. The tubes were centrifuged at 300 g for 5 mins and the supernatant aspirated. The pellet was washed twice by the addition of PBS and centrifugation at 300 g for 5 mins. After the final wash the cell pellet was resuspended in 500 µl PBS and stored on ice for flow cytometry.

3.2.5.2. FLOW CYTOMETRY ANALYSIS

Cells were analyzed on a Epics® XL-MCL flow cytometer (Beckman Coulter, Fullerton, California) using forward and side scatter. This also allowed for elimination of cell debris and aggregates by gating. The same flow cytometer settings were used throughout the study. GFP was excited by an argon laser and fluorescence was detected using a 530/30 nm bandpass filter in the FL1 channel. Ten thousand events were counted for every analysis. Each experiment was carried out in triplicate (appendix). Statistical analysis was perfored using the chi square test $P<0.05$.

3.3. RESULTS

3.3.1. DISTRIBUTION OF THE G1888A SUBSTITUTION AMONG VARIOUS GENOTYPES

Analysis of 519 precore sequences from GeneBank (table 3.4) revealed the presence of the G1888A substitution only in subgenotype A1 isolates except for one genotype C isolate from a Japanese HCC patient (Takahashi *et al*., 1998).

Table 3.4. Distribution of the G1888A mutation in different genotypes of HBV.

Genotype	No of precore sequences analysed	G1888A
A	93	24 (25.8%)
B	97	0
C	103	1 (0.9%)
D	64	0
E	50	0
F	31	0
G	11	0
H	20	0

3.3.2. THE EFFECT OF G1888A ON THE ΔG OF THE ENCAPSIDATION SIGNAL.

To test whether this mutation had any effect on the stability of the encapsidation signal, the free energy level of the encapsidation signal with and without the 1888 G-to-A mutation was calculated using *mfold* (Zuker, 2003). There was a -1º energy change with the introduction of the A at nucleotide position 1888, indicating a slight increase in stability (figure 3.2).

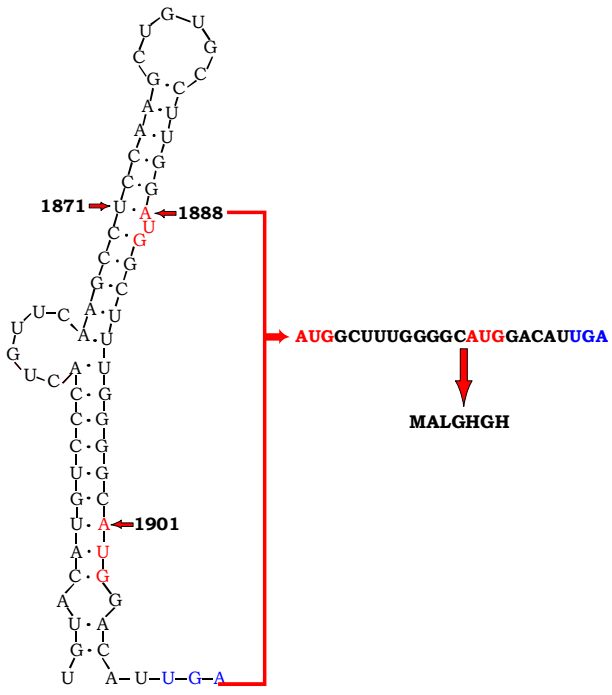


Figure 3.2. The nucleotide sequence and predicted secondary structure of the HBV encapsidation signal of subgenotype A1 and the introduction of a start codon at position 1888 caused by the G-to-A substitution. The translation of the uORF resulting in a 7 amino acid peptide is shown on the right hand side. The introduced A at position 1888 forms a Watson and Crick bond with U1871 resulting in a slightly more stable encapsidation signal.

3.3.3. THE EFFECT OF THE G1888A MUTATION WAS NOT DETECTABLE USING COUPLED TRANSCRIPTION/TRANSLATION EXPERIMENTS *IN VITRO*.

To investigate whether translation of the newly introduced uORF has any effect on translation of the downstream core protein, coupled transcription translation experiments were performed using plasmids; precore1, precore2, precore control, core1 and core2 (table 3.3). The first three plasmids were included to see if the combination of the G1809T/C1812T together with the G1888A had a different effect on core protein translation from the precore mRNA when compared to the wild-type (precore control). There was no evident difference in lanes 3-5 (figure 3.3), indicating that this mutation did not have a measurable effect on translation of the core on the precore mRNA at least in this experimental model. The pregenomic RNA produces the majority of core protein during HBV viral replication. Translation of the core protein from the plasmid expressing only the core protein ('pregenomic RNA') did not differ between plasmid Core1 and Core2 lanes 6 and 7 (figure 3.3), indicating that the system could not detect any difference in core protein translation between the two plasmids.

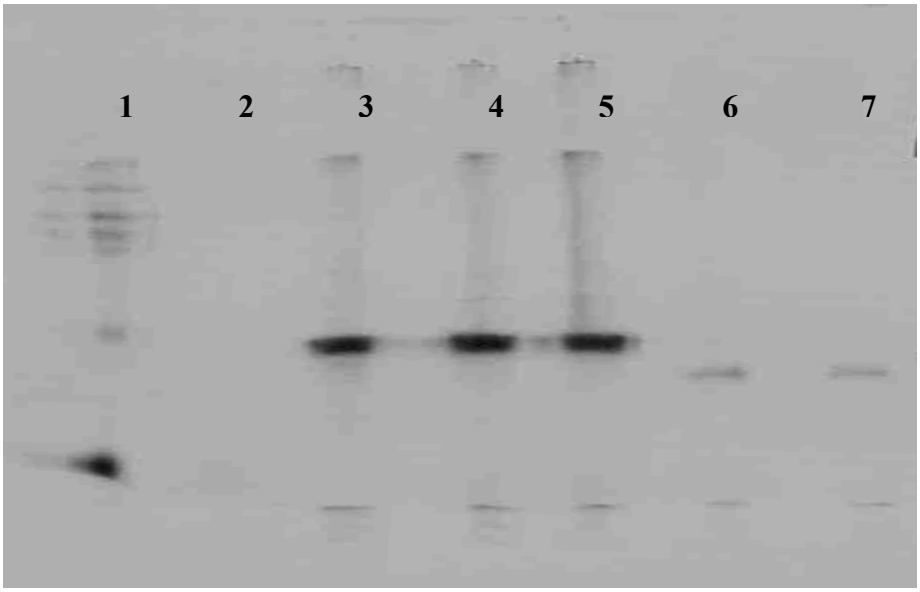


Figure 3.3. *In vitro* transcription/translation of precore and core plasmids. Lane 1 represents the molecular weight marker, lane 2 the negative control, lane 3 precore control, lane 4 precore 2, lane 5 precore 1, lane 6 core 1, lane 7 core 2.

Since coupled transcription/translation could not detect differences in translation between the plasmids, a more sensitive method was devised.

3.3.4. MEASUREMENT OF TRANSLATION USING GFP EXPRESSING VECTOR AND FLOW CYTOMETRY.

A minicistron equivalent to the uORF created by the G1888A substitution was introduced into a wild-type green fluorescent protein expressing plasmid {(pWay19) generously provided by Thomas Hughes, Molecular Motion Lab Montana State University} using primers GFP1+ and GFP1- to produce UpWay19. The reverse primer GFP- contained nucleotide changes that introduced a translation initiation context,

which was the same as that introduced by the G1888A mutation, whereas the forward primer introduced the stop of the uORF. The resulting minicistron, although not the same in aa composition as that introduced by the G1888A mutation, was 7 aa in length and terminated 5 nucleotides from the GFP initiation codon. pWay19 (wild-type GFP) and UpWay19 (uORF + GFP) were transfected into HuH7 cells. Forty-eight hrs post transfection GFP production was quantified at a single cell level using flow cytometry. Experiments were repeated three times. A single experiment is represented in figure 3.4. Figure 3.5 represents the average of 3 experiments. Introduction of this minicistron lead to a 10% decrease in GFP production ($P<0.05$).

Figure 3.4 Flow cytometry analysis of GFP expression. i represent the GFP count of the total cell population. I in i indicate the count from cells expressing GFP i.e those in quadrant 1 of ii. The mean of these count is represented as the amount of GFP produced (top right hand corner of I). The area to the left of I (N) represents cells that do not express GFP i.e those in quadrant 3 of ii. L represents the position of the internal standard used for monitoring variation between flow cytometry analyses. Negative control (A) represents density plots of HuH7 cells transfected with OPTI- MEM, (B) HuH7 cells transfected with pWay19, and (C) HuH7 cells transfected with UpWay19 (uORF + GFP). GFP expression is measured on the y-axis of quadrant 1.

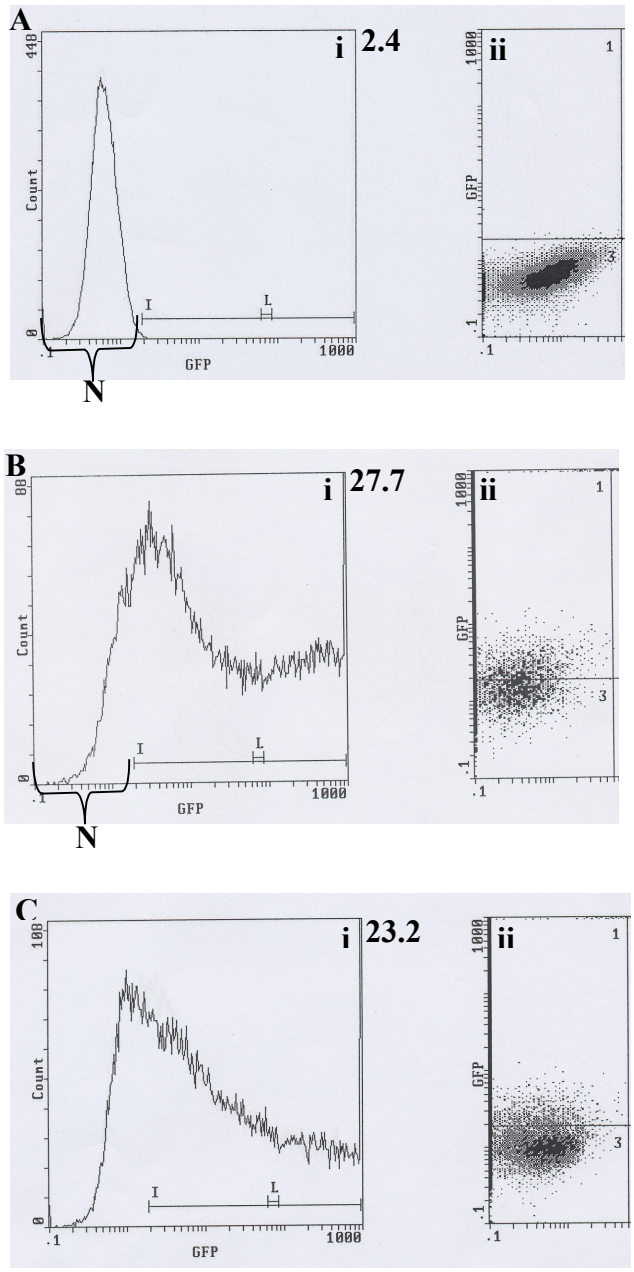


Figure 3.4

A=Negative, **B** =pWay 19, **C**=uORF + GFP

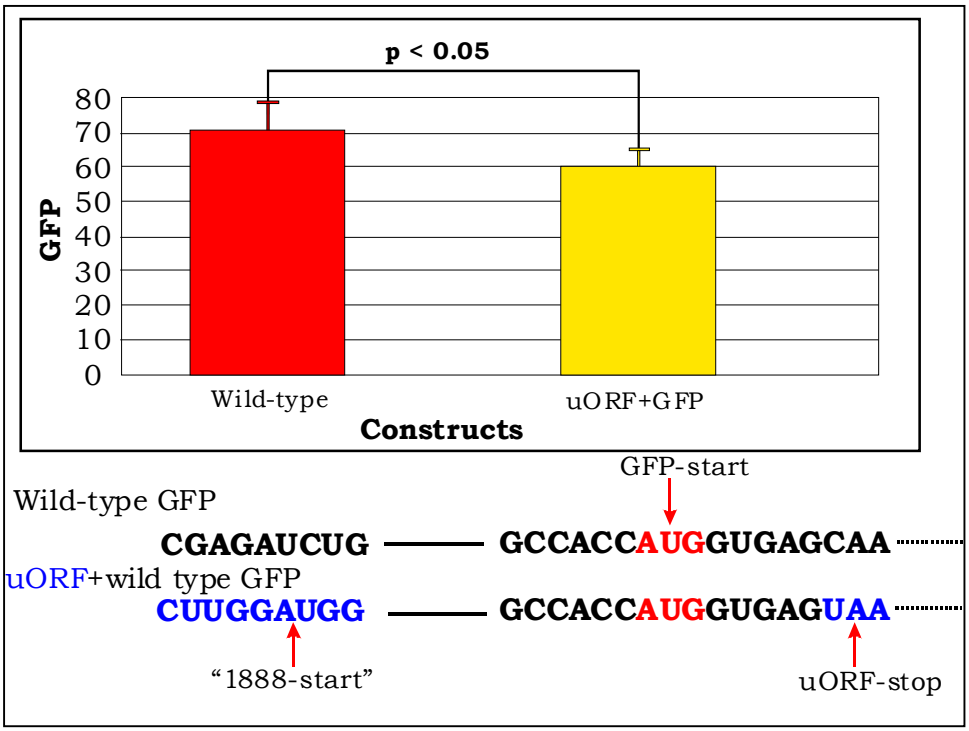


Figure 3.5. Expression of GFP in HuH7 cells transfected with GFP-expressing plasmid pWay19 and UpWay19. GFP expression in percentage is measured in arbitrary units using a flow cytometer and the average of 3 experiments is shown.

The above results indicated that fusion of the core gene to the amino end of a GFP expressing plasmid would result in more sensitive quantification of core protein expression.

3.3.5. G1888A RESULTS IN A DECREASE IN CORE PROTEIN PRODUCTION. THIS COULD BE DETECTED UPON IN-FRAME FUSION OF THE COMPLETE HBV CORE GENE TO THE GFP–ENCODING SEQUENCE.

In order to measure the effect of the uORF on core protein translation, the complete core gene of subgenotype A1 was amplified using primers coreF and coreR. This was then fused in frame on the amino end of a GFP-encoding sequence of the GFP expressing plasmid pWay20 to yield plasmid coreA1. A polyglycine chain (indicated in red on coreR table 3.2) was introduced between the core protein encoding sequence and the GFP-encoding sequence to allow for flexibility of the fusion protein and better detection (Kratz *et al*., 1999). The reverse primer was also designed to eliminate the core stop codon. Core fusion to the amino terminal of pWay20 allowed translation of core/GFP fusion protein to commence from the core AUG since the AUG of the GFP encoding sequence was eliminated. Because the core gene was in-frame with the GFP-encoding sequence the amount of GFP expressed would be proportional to the core gene expression. CoreA2 was generated by site directed mutagenesis on coreA1 with primers S1888F and S1888R. CoreA1 (wild type) and coreA2 (G1888A) were transfected into HuH7 cells. 48hrs post-transfection GFP production was quantified at a single cell level using flow cytometry. Experiments were repeated in triplicate. Figure 3.6 shows flow cytometry analysis of GFP expression of a single experiment using 3 plasmids coreA1 and coreA2 and coreA1+Opt. Figure 3.7 represents the average of all the experiments. Introduction of this minicistron lead to an 18.75 % decrease in GFP production ($P<0.05$).

3.3.6. LEAKY SCANNING ACCOUNTS IN PART, FOR THE DECREASE IN CORE PROTEIN PRODUCTION.

To investigate whether leaky scanning may account for some of the reduction in core protein translation observed, the suboptimal Kozak initiation sequence of the 1888 start was replaced with an optimal Kozak sequence using site directed mutagenesis with primers 1888kozF and 1888KozR on plasmid coreA1 to yield coreA1+Opt. If leaky scanning resulted in reduction of core protein translation an optimal Kozak sequence surrounding the 1888AUG would result in a more drastic reduction of core protein translation. HuH7 cells were transfected with plasmid coreA1+Opt. Forty-eight hrs post transfection GFP expression was quantified using flow cytometry (Figure 3.6). The introduction of an optimal Kozak sequence resulted in a 64.84% reduction in core protein translation ($P<0.05$) (figure 3.7).

Figure 3.6. Flow cytometry analysis of GFP expression. i represent the GFP count of the total cell population. I indicate the count from cells expressing GFP i.e those in quandrant 1 of ii. The mean of these count is represented as the amount of GFP produced (top right hand corner of i). The area to the left of I (N) represents cells that do not express GFP i.e those in quadrant 3 of ii. L represents the position of the internal standard used for monitoring variation between flow cytometry analyses. Wild-type (A) represents density plots of HuH7 cells transfected with coreA1, (B) G1888A transfected with coreA2, and (C) G1888A + Opt transfected with coreA1 + Opt. GFP expression is measured on the y-axis of quadrant 1.

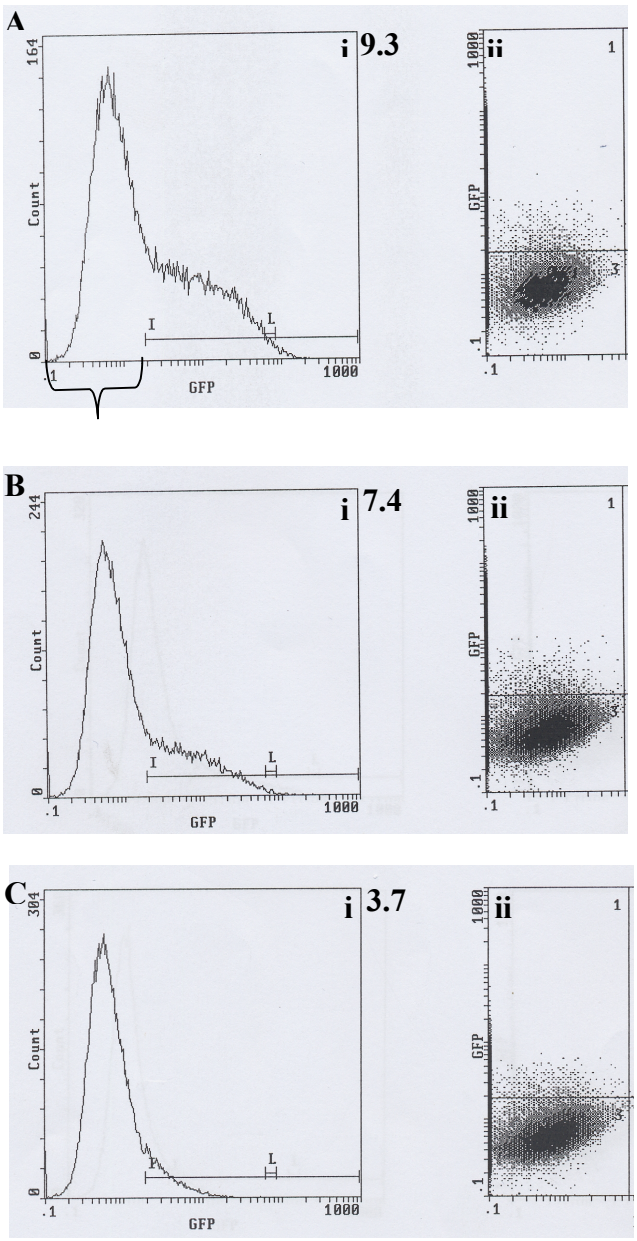


Figure 3.6

A=Wild-type, **B**=G1888A, **C**=G1888A+ Opt

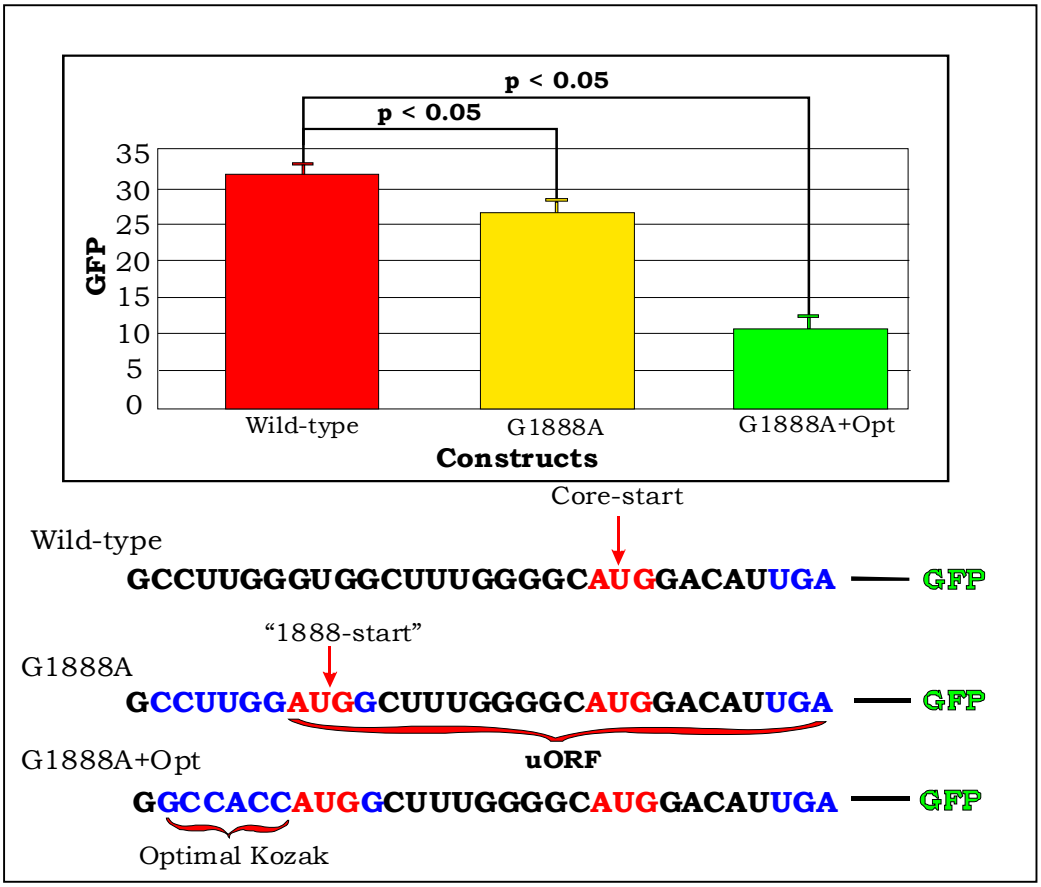


Figure 3.7. Expression of GFP in HuH7 cells transfected with core/GFP-expressing plasmid with and without G1888A substitution measured in arbitrary units. (Opt: optimal Kozak sequence) The experiments were done in triplicate. Wild type (coreA1) = 1888G, coreA2= G1888A. Core A1+Opt= G1888A+Opt.

3.3.7. STALLING OF RIBOSOMES AT THE STOP CODON OF THE UORF MAY INTERFERE WITH TRANSLATION INITIATION AT THE CORE AUG.

To investigate whether translation termination at the stop of the uORF was interfering with initiation at the core AUG, we constructed three plasmids in which the distance between the core start codon and the uORF stop codon was increased. The primer sets and templates used to obtain these plasmids are indicated in tables 3.1 and 3.2. In G1888A(15) the uORF stop was 15 nucleotides from the core start codon, in G1888A(60) it was 60 nucleotides away, and G1888A(100) it was 100 nucleotides away. HuH7 cells were transfected with plasmids G1888A(5), G1888A(15), G1888A(60), and G1888A(100). Forty-eight hrs post transfection GFP expression was quantified using flow cytometry (figure 3.8). By increasing the distance by a minimum of 15 nucleotides, there was a 20% increase in GFP expression. Further elongation of this distance had no significant effect on core protein expression ($P<0.05$) (figure 3.9).

Figure 3.8. Flow cytometry analysis of GFP expression. i represent the GFP count of the total cell population. I indicate the count from cells expressing GFP i.e those in quadrant 1 of ii. The mean of these count is represented as the amount of GFP produced (top right hand corner of I). The area to the left of I (N) represents cells that do not express GFP i.e those in quadrant 3 of ii. L represents the position of the internal standard used for monitoring variation between flow cytometry analyses. CoreA1+Opt was used as a template to minimize interference of ribosomes resulting from leaky scanning of at the 1888 AUG.

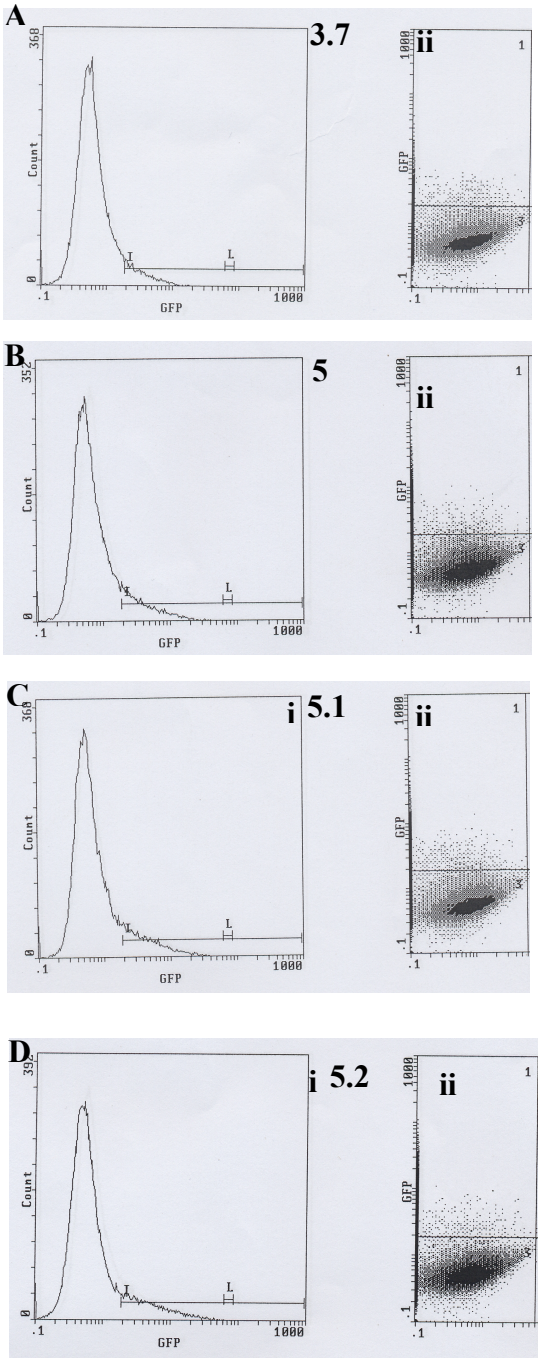


Figure 3.8

A=G1888A(5) (G1888A+Opt), B=G1888A(15), C=G1888A(60),

D=G1888A(100).

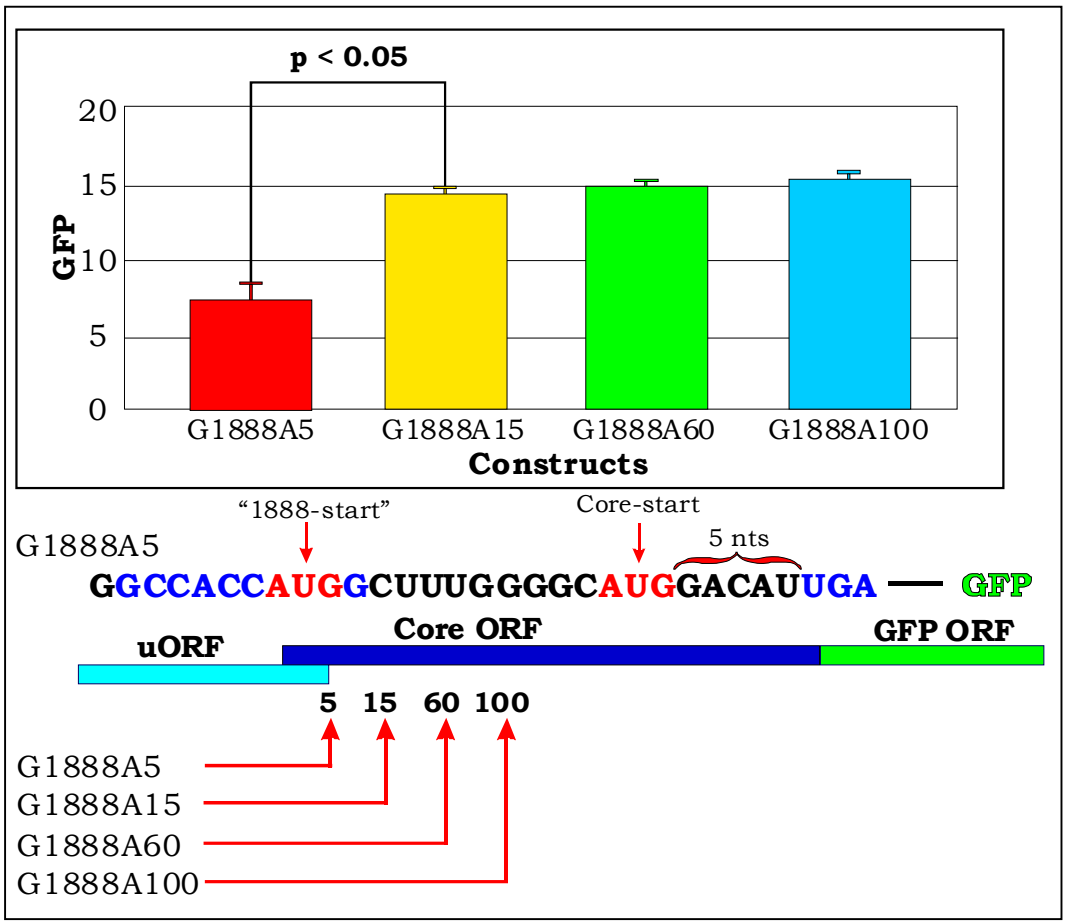


Figure 3.9. Expression of GFP in HuH7 cells transfected with core/GFP-expressing plasmids with G1888A substitution and varying distances of the uORF stop codon from the core AUG. Experiments were carried out in triplicate.

3.4. DISCUSSION

Analysis of 519 precore sequences of HBV isolates from all eight HBV genotypes indicated that the G1888A substitution occurred only in subgenotype A1 isolates. When analysis was limited to 59 subgenotype A1 isolates from GeneBank, the distribution of the G1888A mutation was 64% and 70% for isolates from South Africa and non-South African isolates, respectively. Nucleotide 1871T was the corresponding base pairing nucleotide in all genotypes indicating that there was no constraint at base pairing level for this substitution to be restricted to subgenotype A1. Analysis of substitution that might be linked with the G1888A substitution pointed to the G1809T/C1812T substitution. This has been shown previously to occur only in subgenotype A1 (Ahn *et al*., 2003 and Sugauchi *et al*., 2004) and may lead to a 20% decrease in HBeAg production (Ahn *et al*., 2003). Introduction of the A at nucleotide 1888 of the encapsidation lead to an increase in negative free energy by 1°. This indicates some level of increased stability although it is probably minimal.

Coupled *in vitro* transcription/translation experiments with rabbit reticulocytes were not sensitive enough to detect any differences in the level of the core protein translated from precore and core protein expressing constructs with or without the G1888A mutation (figure 3.3).

Flow cytometry analyses of GFP expressing cells have been shown to have sensitivity similar to other well known reporter systems and this system is even more effective when transfection efficiency or GFP expression levels are low (Ducrest *et al*., 2002).

Therefore this system was used to investigate the effect of the G1888A mutation on translation of the core protein.

Introduction of a minicistron similar to the uORF introduced by the G1888A on a GFP expression plasmid resulted in a 10% decrease in GFP expression (figure 3.5). Given that the AUG of the GFP encoding sequence had an optimal Kozak initiation sequence we expected a more pronounced effect upon examination of the HBV core gene, which is under the control of a sub-optimal initiation context. There was an 18.75% decrease in core-GFP fusion protein production (figure 3.7). This was expected as the AUG introduced by the G1888A mutation had a weak Kozak initiation sequence (figure 3.7). Hence, it exerts a minimal reducing effect by taking up a limited amount 5'-3' migrating ribosomes before they reach the core AUG.

Since the introduced uORF was out-of-frame and overlapped with the core AUG, leaky scanning was the most likely mechanism for the decrease in core protein production. This possibility was investigated by changing the Kozak initiation sequence around the 1888 AUG to an optimal one. By initiating at the 1888 optimal AUG most of the migrating ribosomes will be prevented from reaching the core AUG. This was found to be the case. A 64.84% reduction in core protein production upon introduction of an optimal context around the 1888AUG (figure 3.7) confirmed that leaky scanning was at least in part responsible for core protein reduction.

Next we investigated whether the stop of the uORF was interfering with initiation at the core AUG. This was done by increasing the distance between the core AUG and the stop of the uORF without changing the amino acid sequence of the core protein. It has been shown by ribonuclease protection experiments that the distance between the ribosome's leading 3' edge and the AUG recognition centre is 12-15 nucleotides (Kozak, 1977). If the termination codon of the uORF was interfering with initiation at the core AUG increasing the distance between them to a minimum of 15 nucleotides would alleviate the steric hindrance effect leading to an increase in core protein production. A further increase should have no additional significant effect. There was a 40% increase in core protein translation when this distance was increased to 15 nucleotides (figure 3.8). Further elongation of the distance between the core start and uORF stop had a minimal effect on core protein translation. Stalling of ribosomes at the stop codon of uORFs has been reported in the past (Cao and Geballe, 1996a, Cao and Geballe, 1996b). This can be attributed to the inability of the final tRNA to detach from mRNA. These tRNAs together with the ribosomal complex form a barrier preventing scanning of 40s ribosomes. If stalling were to occur at the stop of the introduced uORF this would interfere with initiation at the core AUG through a steric hindrance mechanism because of the close proximity between them.

The HBeAg functions as immuno-regulatory protein. This is achieved by down regulating the host T cell response to the intracellular nucleocapsid (HBcAg) via a selection of immune tolerance inducing mechanisms, resulting in a less severe immune reaction to the infecting viral particles and hence minor liver injury (Chen *et al*., 2004).

A decrease in HBeAg production by any of the variants of the 1809/1811/1812 substitutions (Ahn *et al*., 2003) may result to a more severe immune response, liver damage and acute state of hepatitis. The G1888A mutation, by reducing the amount of HBcAg produced may compensate or counterbalance the decrease in immuno-suppressive effect as a result of a reduction in HBeAg production by any of the above-mentioned variants.

The core protein together with the viral polymerase and viral pregenome comprise the core particle, which is the location for HBV replication. A decrease in core protein production will lead to a decrease in core particle formation and this reduces replication levels. This together with changes in the DNA binding cleft of the polymerase found in subgenotype A1 may account for the low levels of replication seen in individuals infected with this strain.

3.5. CONCLUSION

The G1888A substitution occurs only in subgenotype A1 isolates. The introduction of the uORF resulting from the G1888A substitution into either a GFP expressing plasmid or a core/GFP fusion plasmid can reduce the expression of the protein coded for downstream. This reduction was further enhanced when the context of the Kozak initiation sequence preceding the newly introduced 1888 AUG start codon was optimal, intimating that leaky scanning at the 1888 AUG may be responsible for the reduction. Increasing the distance between the core start and stop of the uORF by a minimum of 15 nucleotides resulted in an increase in core protein translation indicating that the stop of the uORF was interfering with translation of the core.

This minimal decrease in core protein translation immediately after infection may lead to fewer core antigen epitopes being recognised. This might compensate for the decreased production of the immuno-suppressive HBeAg as a result of the G1809T/ C1812T (the most common variant of the 1809/1811/1812 substitutions in genotype A1) resulting in a less severe immune response against the virus and hence less liver injury. This might lead to the development of a chronic or asymptomatic state of infection predisposing the individual to HCC.

CHAPTER 4

ANALYSIS OF GENOTYPE D ISOLATES FROM SOUTH AFRICA

SUMMARY

HBV genotype D is the second most abundant genotype in South Africa. The complete genomes of six genotype D HBV isolates have been amplified from asymptomatic carriers of HBV from South Africa. Comparison of these complete genomes with complete genotype D genomes from GeneBank revealed the presence of four subgenotypes (D1-D4) based on significant bootstrap values of greater than 75% at the level of the complete genome. All six genotype D isolates from South Africa clustered together in a separate minor clade (Dsa) within subgenotype D3. Separation into 4 subgenotypes was well supported at the level of the complete genome, polymerase, complete S gene and precore gene. Amino acids specific to D3 were identified in the polymerase region. Some amino acid changes within this region were only shared between Dsa isolates and isolates from other non-D genotypes. Specific nucleotide changes where also identified in D3 within the enhancer I/ X promoter and the S2 promoter. The low intragroup divergence between Dsa members suggests a recent introduction of this genotype in South Africa.

4.1. RESULTS

Six of the twenty three complete genomes sequenced in the present study belonged to genotype D (figure 4.1). These sequences have been deposited in GeneBank (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed) (GeneBamk no: AY233291-AY233296). Subgenomic PCR and direct sequencing of the amplicons confirmed that the clones sequenced were representative of the major HBV strains in the serum.

4.1.1. PHYLOGENTIC ANALYSIS

The length of the complete genomes of genotype D isolates sequenced in the present study was 3182 bp in length. The serological subtype of all Dsa isolates was deduced from the sequence to be *ayw2*. All Dsa isolates had a 6 bp deletion between nucleotide 2356 and 2363 of the core gene. The 33 bp deletion at the beginning of the pre-S1, characteristic of genotype D was also present in the Dsa isolates.

.

All genotype D isolates clustered together into four separate subgenotypes (D1-D4) upon phylogenetic analysis of the complete genome, polymerase, complete S gene and precore gene (figure 4.1 and 4.4) (Kramvis *et al*., 2005). This separation was most supported at the level of the complete genome with bootstrap values of 76-99. Genotype D isolates could not be separated into four subgenotypes in the X gene. All South African isolates clustered together in a minor clade Dsa. This was most supported at the level of the core gene (figure 4.4). The mean nucleotide divergence between Dsa and the rest of the genotype D isolates at the level of the complete genome was 3.49%.

This was also similar when D3 was compared with D1/D2 3.71% (table 4.1). The intagroup divergence of D3 members was 1.31% (table 4.1).

Figure 4.1-4.5 Dendrograms of complete HBV genomes and the four ORFs of genotype D isolates together with representatives of the remaining seven genotypes. Geneotype D could be separated into four subgenotypes (D1-D4) at the complete genome level, in the polymerase, complete S gene and precore gene. Dsa clusters together in all ORFs. This cluster is most supported at the level of the precore gene bootstrap value 99. Samples in red are South African isolates from the present study. All genotype D samples from the present study were isolated from asymptomatic HBV carriers. SA= South Africa.

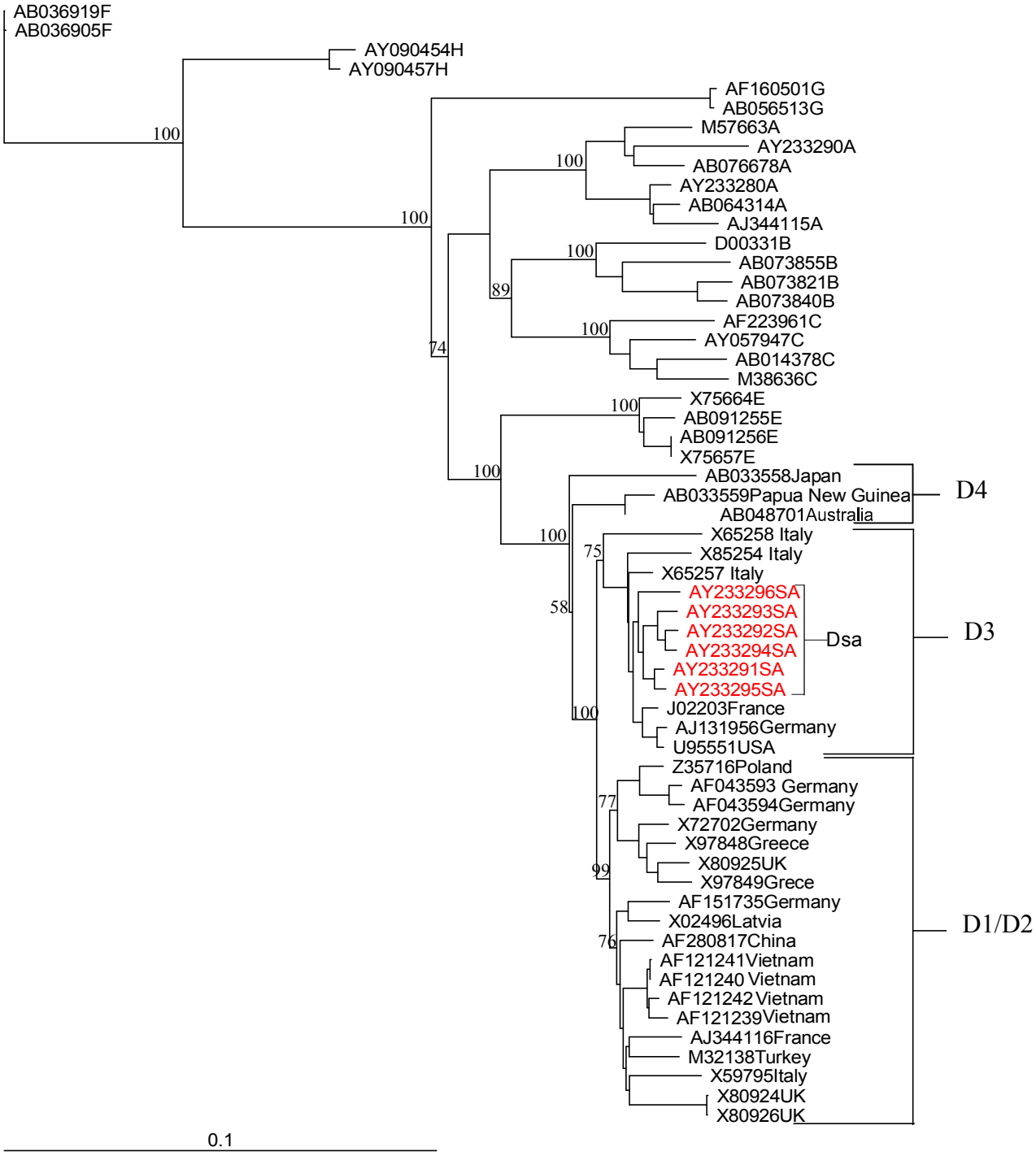


Figure 4.1. Complete Genome.

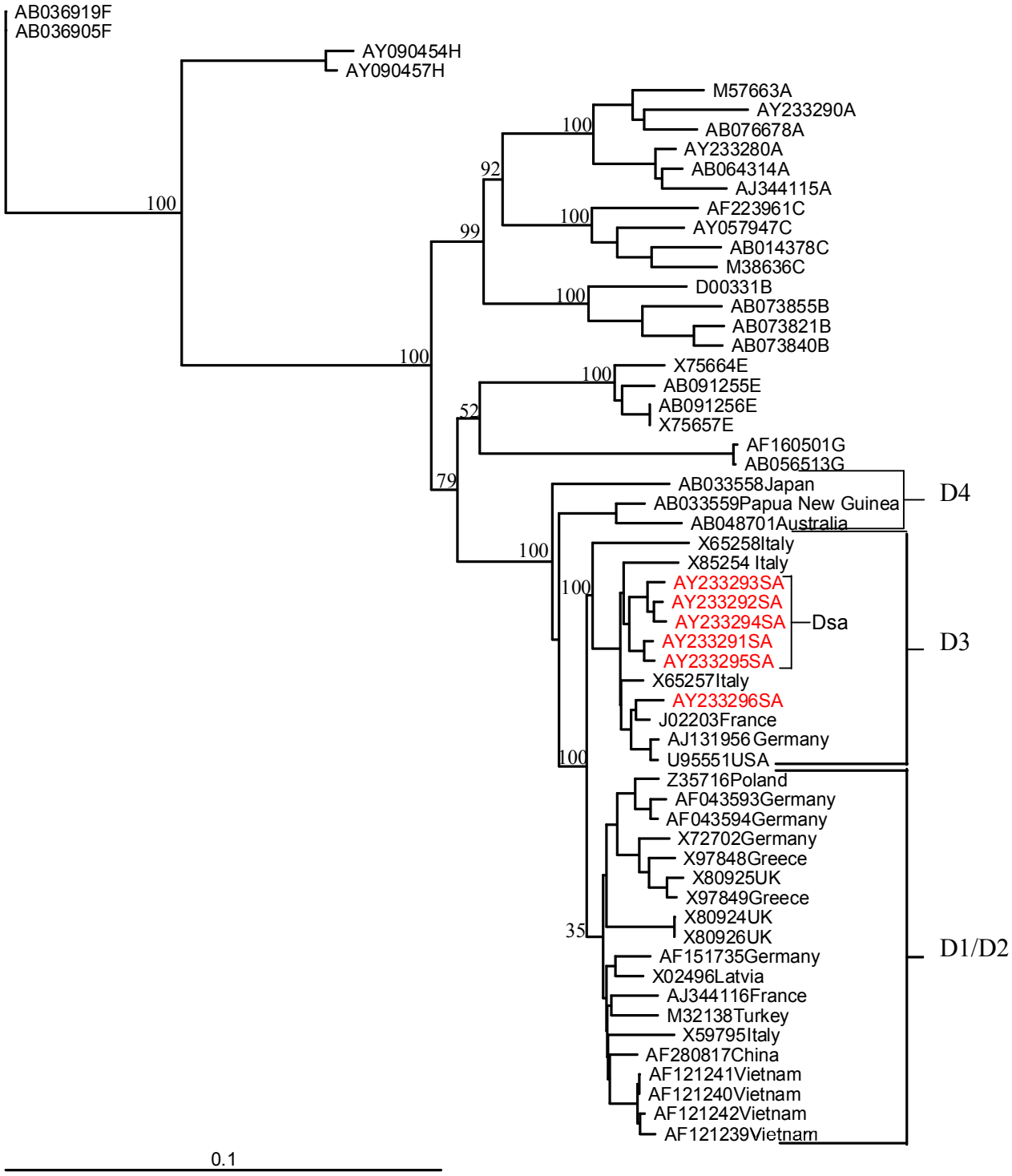


Figure 4.2. Polymerase gene

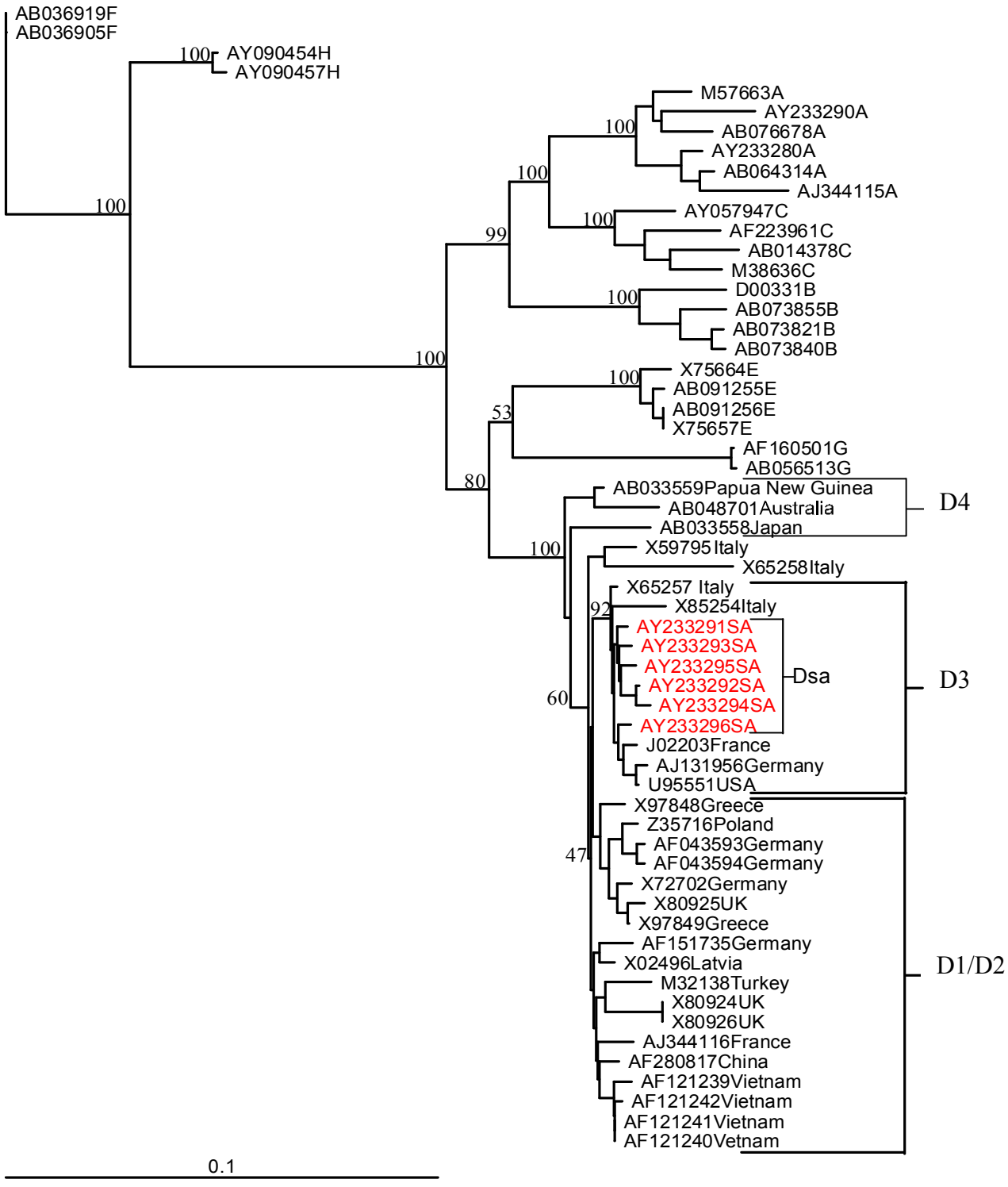


Figure 4.3. Complete S gene

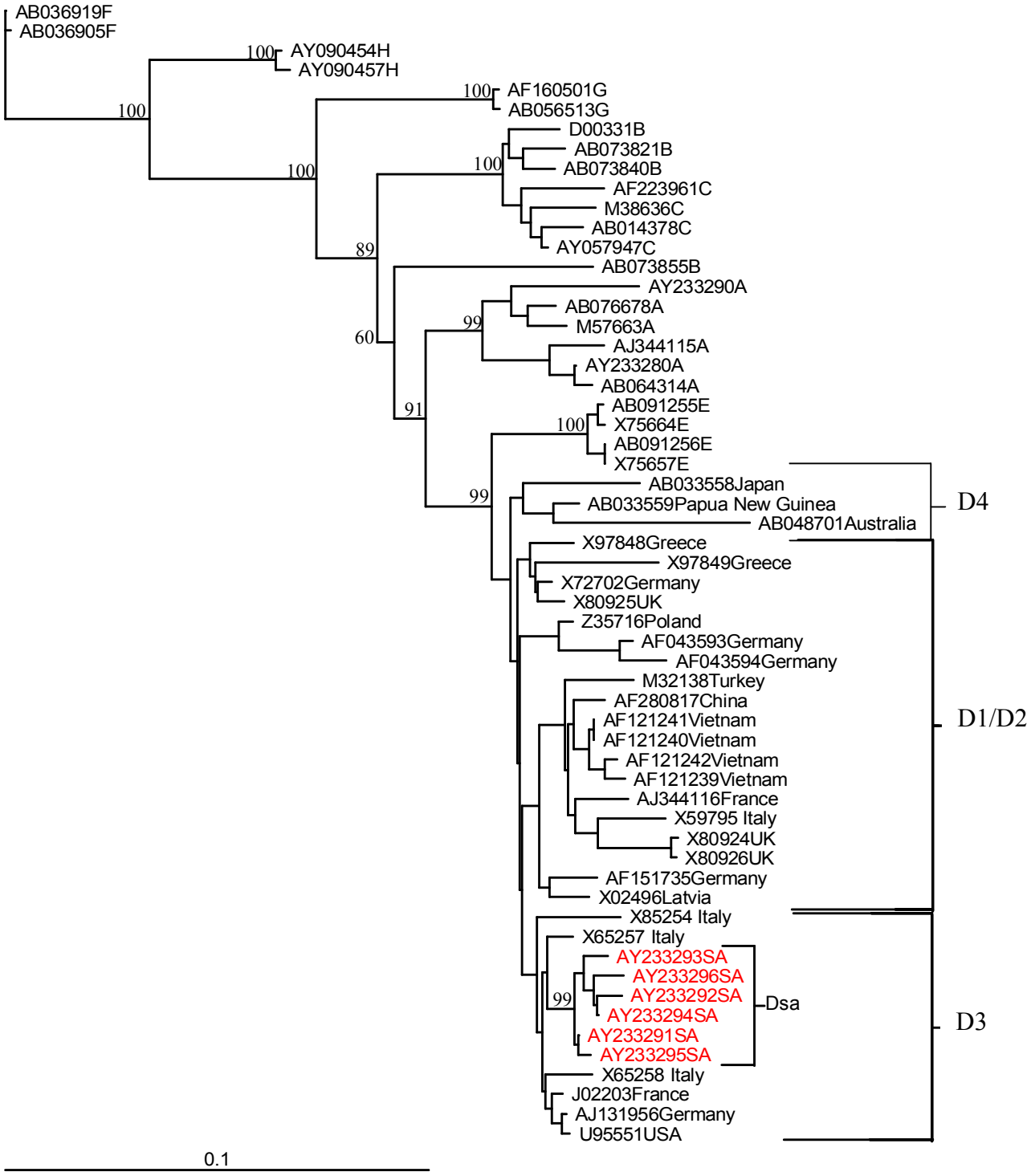


Figure 4.4. Precore/core gene

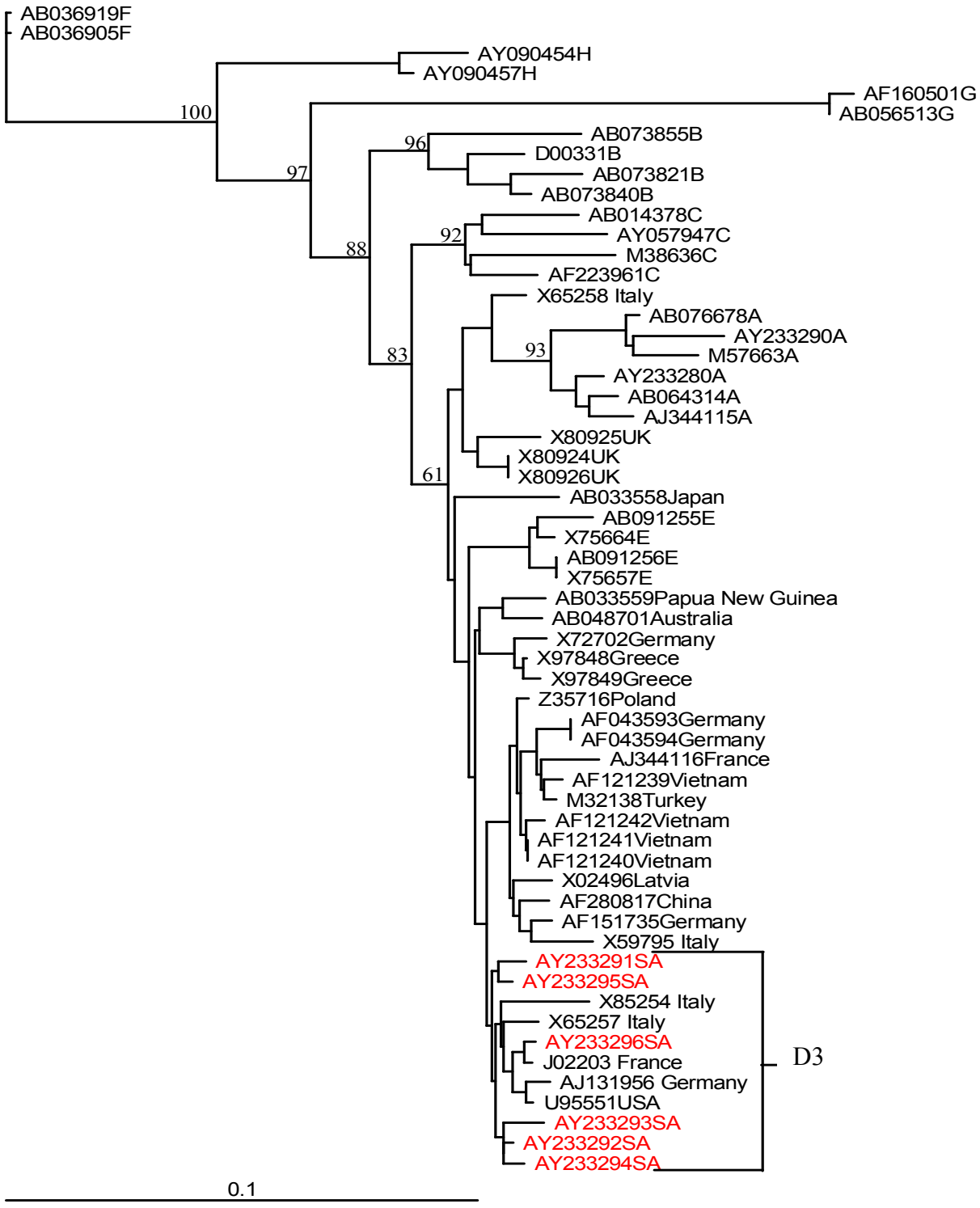


Figure 4.5. X gene

4.1.2. COMPARISON OF AMINO ACID SEQUENCES OF D3 TO THOSE OF OTHER GENOTYPE D MEMBERS AND OTHER GENOTYPES

Figure 4.6 provides a comparison of the translated sequences of the six Dsa isolates, with sequences obtained from GeneBank. The amino acids found only in D3 are shown in red, those found in D3 and in other non-D genotypes but only in a minority of the remainder genotype D isolates are shown in bold. Those shared only between Dsa and other non-D genotypes are shaded in light green. Leu^{42} of the pre-S2 together with, Phe^{249}, and Leu^{343} in the spacer of the polymerase were unique to D3. Val^{463} of the polymerase spacer was only shared between Dsa and genotype H. Ile^{69} of the polymerase priming region was shared between Dsa and genotypes E and G. Ala^{711} and Ile^{712} was shared between Dsa and all other non-D genotypes. Pro^{478} in the polymerase RT domain and Val^{102} in the core were shared by D3 and genotype E. Ile^{88} in the X region was shared with genotype A and E.

By corollary, D1, D2, and D4 genotype D isolates also have signature amino acids that distinguish them from D3 and all other genotypes. They are circled in figure 4.6. They include Ile^{249} and Tyr^{279} of the polymerase spacer: Phe^{470} of the RT domain of the polymerase; Val^{711} and Met^{712} of the RNaseH of the polymerase; and Ser^{46} and Phe^{88} of the X protein.

4.1.3. COMPARISON OF THE NUCLEOTIDE SEQUENCES OF *CIS*-ACTING ELEMENTS OF D3 TO THOSE OF OTHER GENOTYPE D MEMBERS AND OTHER GENOTYPES

The sequences of the *cis*-acting elements of the six Dsa isolates sequenced in the present study were compared with sequences obtained from GenBank (figure 4.7). Nucleotides unique to D3 are shaded in red, those found in D3 and other non-D genotypes are in bold, whereas those shared mostly between Dsa and other non-D genotypes are shaded in light green. Table 4.2 summarizes the mutations in the *cis*-acting elements characteristic of D3 and the functional elements that are affected by these mutations. A^{961} of the enhancer I region was particular to D3 except for 2 isolates of the remainding genotype D members. T^{3012} of the S2 promoter was only present in subgenotype D3 isolates. C^{1029} and $C/A^{1217/1221}$ were shared by subgenotype D3 isolates and those from genotypes A, B, C, and E. Nucleotide substitutions particular to D1, D2, and D4 are circled (figure 4.7). They include C^{1324}, $G^{1332,}$ and T^{1347} of the enhancer I/X promoter, $T^{1509}$and T^{1635} of the core promoter/ enhancer II region and A^{3012} and T^{3102} of the S2 promoter.

Table 4.1. Mean nucleotide divergence (%) of complete genome and individual ORF sequences of HBV genotype D obtained using DAMBE

	Intragroup					Intergroup		
	Dsa	D3	Genotype D	D^a	D^b	Dsa vs D^a	D3 vs D^b	Dsa vs D3-Dsa
Complete Genome	1.05±0.60¶ (0.51-1.77)	1.31±1.23 (0.38-2.02)	3.16±1.23 (0.03-5.91)	3.12±1.36 (0.03-5.91	3.07±1.36 (0.03-5.91)	3.49±0.84 (1.23-5.31)	3.73±0.58 (2.78-5.31)	1.77±0.20 (1.23-2.00)
Polymerase	1.07±0.63 (0.44-1.77)	1.26±0.61 (0.36-1.94)	2.92±1.18 (0.04-5.48)	2.86±1.25 (0.04-5.48)	2.76±1.32 (0.04-5.46)	3.28±0.78 (0.97-4.95)	3.46±0.52 (2.54-4.96)	1.66±0.23 (0.97-1.94)
X Gene	1.05±0.57 (0.65-1.94)	1.19±0.59 (0.43-2.37)	2.40±1.04 (0.00-4.95)	2.46±1.13 (0.00-4.95)	2.45±1.22 (0.00-4.95)	2.41±0.73 (0.43-4.52)	2.58±0.63 (1.51-4.52)	1.48±0.48 (0.43-2.37)
Precore/Core	.90±0.56 (0.31-1.73)	1.33±0.78 90.31-2.68)	3.80±1.75 (0.00-8.63)	3.83±1.78 (0.00-8.63)	3.95±1.85 (0.00-8.63)	4.12±1.42 (1.26-8.63)	4.14±1.31 (1.89-7.87)	1.99±0.42 (1.26-2.68)
Pre-S1/2	0.52±0.31 (0.12-.85)	0.67±0.38 (0.21-1.27)	2.63±1.38 (0.00-7.19)	2.75±1.48 (0.00-7.19)	2.73±1.58 (0.00-7.19)	2.58±0.98 (0.42-5.29)	2.89±0.84 (1.69-5.71)	.92±0.29 (0.42-1.27)
HBsAg	0.77±0.46 (0.59-1.62)	0.88±0.46 (0.44-1.91)	1.52±0.60 (0.00-3.08)	1.51±0.62 (0.00-3.08)	1.48±0.64 (0.00-3.08)	1.64±0.43 (0.59-3.08)	1.71±0.47 (0.88-3.08)	1.11±0.34 (0.59-1.9)

http://web.hku.hk/-xxia/software/software.htm. The sequences compared are those included in phylogenetic analysis (twenty eight genotype D sequences from GeneBank and six south African genotype D sequences from present study).

¶ The mean nucleotide divergence (%)± and the range in parenthesis.

D-Dsa represents none south African isolates present in D3

D^a represents all genotype D isolates except those from South Africa.

D^b (D1, D2 and D4) represents all genotype D isolates except those in D3.

Figure 4.6. Comparison of amino acid residues of S, polymerase and X ORFs of subgenotype D3 isolates with amino acids residues of the remaining genotype D subgenotypes and other HBV genotypes. Dots indicate amino acid identity. Numbering according to *adw2* genome (GeneBank accession #AY233276). South African isolates sequenced in the present study are shown shaded in red the remaining sequences were obtained from GeneBank. For non-genotype D isolates the representative amino acid was obtained 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 D3 and in other non-D genotypes but not in D1/D2/D4 are shown in light yellow, those unique to D3 are shaded in red. Amino acids unique to D1/D2/D4 are circled. Amino acids predominant to Dsa when analysis is restricted to genotype D isolates are shaded in light green.

	Genotype/ GenBank#	S REGION	POLYMERASE														CORE GENE	X REGION		
		PRE-S2	Priming region		SPACER					RT						RNaseH				
	Codon #:	42	69	120	249	270	279	306	343	369	463	470	478	611	626	711/712	747	102	33	46	88
		(I)	(V)	(N)	(I)	L	(Y)	(L)	S	A	L	(F)	Q	D	V	(V/M)	(L)	V	P	(S)	(F)
D1/D2/D4	AB033558	.	.	.	V	F	.	F	.	.	.	.	.	.	.	A.	.	T	.	P	I
	AB033559	.	.	.	V	F	.	.	.	.	.	I	.	N	.	.	.	T	.	P	I
	AB048701	T	.	.	V	.	.	.	.	.	.	I	.	.	.	.	.	T	.	P	.
	AF043593	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
	AF043594	.	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
	AF121241	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
	AF121242	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
	AF121239	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
	AF121240	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
	AF151735	.	.	.	.	R	.	.	.	.	.	.	.	.	I	.	.	G	.	.	.
	AF280817	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
	AJ344116	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
	M32138	.	.	K	.	.	H	.	.	.	.	.	.	.	.	.	.	G	.	.	.
	X02496	.	.	.	.	V	.	.	.	.	.	.	.	.	.	.	.	G	S	.	.
	X72702	.	.	.	.	.	.	.	.	.	V	.	.	.	I	.I	.	.	.	P	.
	X80924	.	.	.	.	.	H	.	.	.	.	.	.	.	.	.	.	G	.	.	M
	X80925	.	.	K	.	F	.	.	.	.	.	L	.	.	.	.	.	.	.	.	I
	X80926	.	.	.	.	.	H	.	.	.	.	.	.	.	.	.	.	G	.	.	M
	X97848	.	.	.	.	.	.	.	.	.	.	I	P	.	.	.	.	.	.	.	.
	X97849	.	.	.	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.	.
	Z35716	.	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	X59795	.	.	M	T	R	.	F	.	.	.	.	.	.	.	.	.	G	.	.	.
D3	AY233292	L	I	K	F	F	H	F	L	S	V	L	P	E	I	A/I	I	.	S	P	I
	AY233293	.	I	K	F	F	H	F	.	S	V	L	P	E	I	A/I	I	.	S	P	I
	AY233291	L	I	.	F	F	H	F	L	.	V	L	P	E	I	A/I	.	.	S	P	.
	AY233295	L	I	.	F	F	H	F	L	S	V	L	P	E	I	A/I	.	.	S	P	I
	AY233294	L	I	K	F	F	H	F	L	S	V	L	P	E		A/I	I	.	S	P	I
	AY233296	L	I	K	V	F	H	F	L	S	.	L	P	E	I		I	.	S	P	I
	J02203	L	.	K	F	F	H	F	L	S	.	L	P	E	I	..	I	.	S	P	I
	AJ131956	L	.	K	F	F	H	F	L	S	.	L	P	E	I	..	I	.	S	.	M
	U95551	L	.	K	F	F	H	F	L	S	.	L	P	E	I	..	I	.	S	P	M
	X65258	L	.	K	T	R	.	.	L	.	.	N	.	E	.	G.	.	.	.	.	I
	X85254	L	.	N	F	F	H	.	L	S	.	L	P	E	I	A.	I	.	S	P	I
	X65257	L	.	K	F	F	H	.	L	S	.	L	P	E	.	A.	I	.	S	P	I
Non-D	GenotypeA	.	.	./T	V/A	S	H	F	.	.	.	H/N	.	.	.	A/I	I	N	.	P	I
	GenotypeB	.	.	.	V	C	H	.	.	.	.	I	.	E	.	A/I	I	S	.	P	N
	GenotypeC	.	.	.	V	S	H	./F	.	.	.	I	.	E	.	A/I	I	S	./S	P	V
	GenotypeE	.	I	I	V	I	H	F	.	.	.	I	P	E	.	A/I	I	.	S	P	I
	GenotypeF	T	.	K	V	./I	H	.	.	.	.	H	.	.	.	A/I	I	N	./S	P	S/N
	GenotypeG	.	I	.	V	F	H	.	.	.	.	L	.	E	.	A/I	I	N	.	P	H
	GenotypeH	T	.	K	V	.	H	.	.	S	V	Y	.	E	.	A/I	I	N	S	p	S/N

Figure 4.6.

Figure 4.7. Comparison of the nucleic acid sequences of the *cis*-acting elements of D1 isolates with the remaining genotype D isolates and other HBV genotypes. Dots indicate amino acid identity. Nucleotide 1 denotes the nucleotide position of HBV *adw2* genome (GeneBank accession # AY233276) where the *Eco*RI cleavage site is at position 1. South African isolates sequenced in the present study are in red and all other sequences were obtained from GeneBank. For non-genotype D isolates the representative nucleotide sequence was obtained by aligning 10 sequences of each genotype from geographically distinct regions and the consensus nucleic acid deduced if it occurred in 60% or more of the sequences. Nucleotides found only in D3 and in other non-D genotypes but not in D1/D2/D4 are shown in light yellow, those unique to D3 are in red. Nucleotide sequences unique to D1/D2/D4 are circled. Nucleotides found only in Dsa and other non-genotype D isolates are shaded in light green. # represents mutations.

	Genotype/ GenBank#	Enhancer I / X Promoter 950-1350							Core Promoter/Enhancer II					S1 Promoter 2716-2806		S2 Promoter 2999-3219		
	Nucleotide #:	961	987	1029	1217/1221	1324	1332	1347	1470	1509	1635	1762/1764	1766	2738	2744	3012	3075	3102
	CONSENSUS	G	C	T	T/G	C	G	T	C	T	T	A/G	C	C	T	A	C	T
D1/D2/D4	AB033558	.	A	.	C/.	.	T	C	.	C	A	./.	.	.	C	G	T	.
	AB033559	.	G	C	..	.	T	C	.	C	A	./.	.	T	C	G	T	.
	AB048701	.	G	C	..	.	T	C	.	C	.	./.	.	T	C	G	.	.
	AF043593	.	.	.	..	.	.	.	.	.	.	T/A	.	.	.	.	.	.
	AF043594	.	.	.	..	.	.	.	.	.	.	T/A	.	.	.	.	.	.
	AF121241	.	.	.	..	.	.	.	.	.	.	./.	.	.	.	.	.	.
	AF121242	.	.	.	..	.	.	.	.	.	.	./.	.	.	.	.	.	.
	AF121239	.	.	.	..	.	.	.	.	.	.	T/A	.	.	.	.	.	.
	AF121240	.	.	.	..	.	.	.	.	.	.	./.	.	.	.	.	.	.
	AF151735	A	.	.	..	.	.	.	.	.	.	./T	G	.	.	.	.	.
	AF280817	.	.	.	..	.	.	.	.	.	.	./.	.	.	.	.	.	.
	AJ344116	.	.	.	..	.	.	.	.	.	.	T/A	.	.	.	.	.	.
	M32138	.	.	.	..	.	.	.	.	.	.	T/A	.	.	.	.	.	C
	X02496	.	.	.	..	.	.	.	T	.	.	./.	.	.	.	.	G	.
	X72702	A	.	.	..	.	.	.	.	C	.	./.	.	.	.	.	.	.
	X80924	.	.	C	..	.	.	.	.	.	A	./.	.	.	.	.	.	C
	X80925	.	.	.	..	.	.	.	.	.	A	./.	.	.	.	.	T	.
	X80926	.	.	C	..	.	.	.	.	.	A	./.	.	.	.	.	.	C
	X97848	.	.	.	..	.	T	.	.	.	.	./.	.	.	.	.	.	.
	X97849	.	.	.	..	.	.	.	.	.	.	./.	.	.	.	.	.	.
	Z35716	.	A	.	..	.	.	.	.	.	.	./.	.	.	.	.	.	.
	X59795	.	.	.	..	.	.	C	.	.	.	./T	G	.	.	.	.	.
D3	AY233292	**A**	**A**	**C**	**C/A**	**A**	**T**	**C**	**T**	**C**	**A**	**./.**	**.**	**T**	**C**	**T**	**T**	**C**
	AY233293	**A**	**A**	**C**	**C/A**	**A**	**T**	**C**	**T**	**C**	**A**	**./.**	**.**	**T**	**C**	**T**	**T**	**C**
	AY233291	**A**	**A**	**C**	**C/A**	**.**	**T**	**C**	**T**	**C**	**.**	**./.**	**.**	**T**	**C**	**T**	**T**	**C**
	AY233295	**A**	**A**	**C**	**C/A**	**.**	**T**	**C**	**T**	**C**	**A**	**./.**	**.**	**T**	**C**	**T**	**T**	**C**
	AY233294	**.**	**A**	**C**	**C/A**	**A**	**T**	**C**	**T**	**C**	**A**	**./.**	**.**	**T**	**C**	**T**	**T**	**C**
	AY233296	**A**	**A**	**.**	**./.**	**A**	**T**	**C**	**T**	**C**	**A**	**./.**	**.**	**T**	**C**	**G**	**T**	**C**
	X65257	**.**	**A**	**.**	**C/.**	**A**	**T**	**C**	**T**	**C**	**A**	**./.**	**.**	**T**	**C**	**T**	**T**	**C**
	X65258	**.**	**A**	**C**	**G/.**	**.**	**.**	**.**	**.**	**.**	**A**	**./.**	**.**	**T**	**C**	**.**	**.**	**.**
	X85254	**A**	**G**	**.**	**C/.**	**A**	**T**	**C**	**T**	**C**	**A**	**./.**	**.**	**T**	**C**	**T**	**T**	**C**
	J02203	**A**	**A**	**.**	**..**	**A**	**T**	**C**	**T**	**C**	**A**	**./.**	**.**	**T**	**C**	**T**	**T**	**C**
	AJ131956	**A**	**A**	**.**	**..**	**A**	**T**	**C**	**T**	**.**	**A**	**./.**	**.**	**T**	**C**	**T**	**T**	**C**
	U95551	**A**	**A**	**.**	**..**	**A**	**T**	**C**	**T**	**C**	**A**	**./.**	**.**	**T**	**C**	**T**	**T**	**C**
Non-D	Genotype A	.	A	.	C/A/C	A	T	C	.	C	A	#	#	.	.	G	A	C
	Genotype B	.	A/T	C/.	C/A	A	T	C/G	.	C	A	#	#	G/T	.	G	T	C
	Genotype C	.	./A	.	C/A/T	A	C/T	C/.	.	C	G	#	#	T	.	G	./A	C
	Genotype E	.	A	C/.	C/A	./A	A	.	T	C	A	#	#	.	C	G	A	C
	Genotype F	.	T	C/.	C/T	A	A	C	./T	C	A	#	#	./A	./C	G	./A	C
	Genotype G	.	.	C/.	C/C	A	T	C	.	C	C	#	#	.	.	G	T	C
	Genotype H	.	./T	.	C/T	A	T	C	T	C	A	#	#	T	./C	G	.	C

Figure 4.7.

Table 4.2. Mutations within the *cis*-regulatory elements found predominantly in D3.

***cis*-acting element**	**Mutation**	**Functional element affected**
Enhancer 1/X	G960A	5'modulator element
	T1029C	C/EBP binding site
	G1221A	NF1
S1 promoter	A3012T	region A

*Those in grey occurred predominantly in Dsa when analysis was restricted to genotype D isolates. Numbering of nucleic acids according to *adw2* genotype [GeneBank accession number AY2333276] and functional domain according to Moolla *et al*., (2002), Patel *et al*., 1989.

4.1.4. ANALYSIS OF THE PRE-S1 REGION OF GENOTYPE D ISOLATES FROM GENEBANK.

The preS1 region of 76 genotype D isolates was analyzed phylogenetically (figure 4.8). This region was chosen because it was the section of the genome with the largest number of sequences deposited in GeneBank. Separation into subgenotype D1 to D4 was evident. South African isolates from previous studies (Hardie & Williamson, 1997) clustered with those from the present study. Samples from different geographical locations clustered together indicating a low degree of variability at the pre-S1 the same population.

Figure 4.8. Phylogram of the pre-S1 region of 76 Genotype D isolates from GeneBank. Separation into D1, D2, D3 and D4 is evident. South African isolates from previous studies are in Violet. Those from the present study are in red. The geographical origin of the genotype D isolates is also indicated.

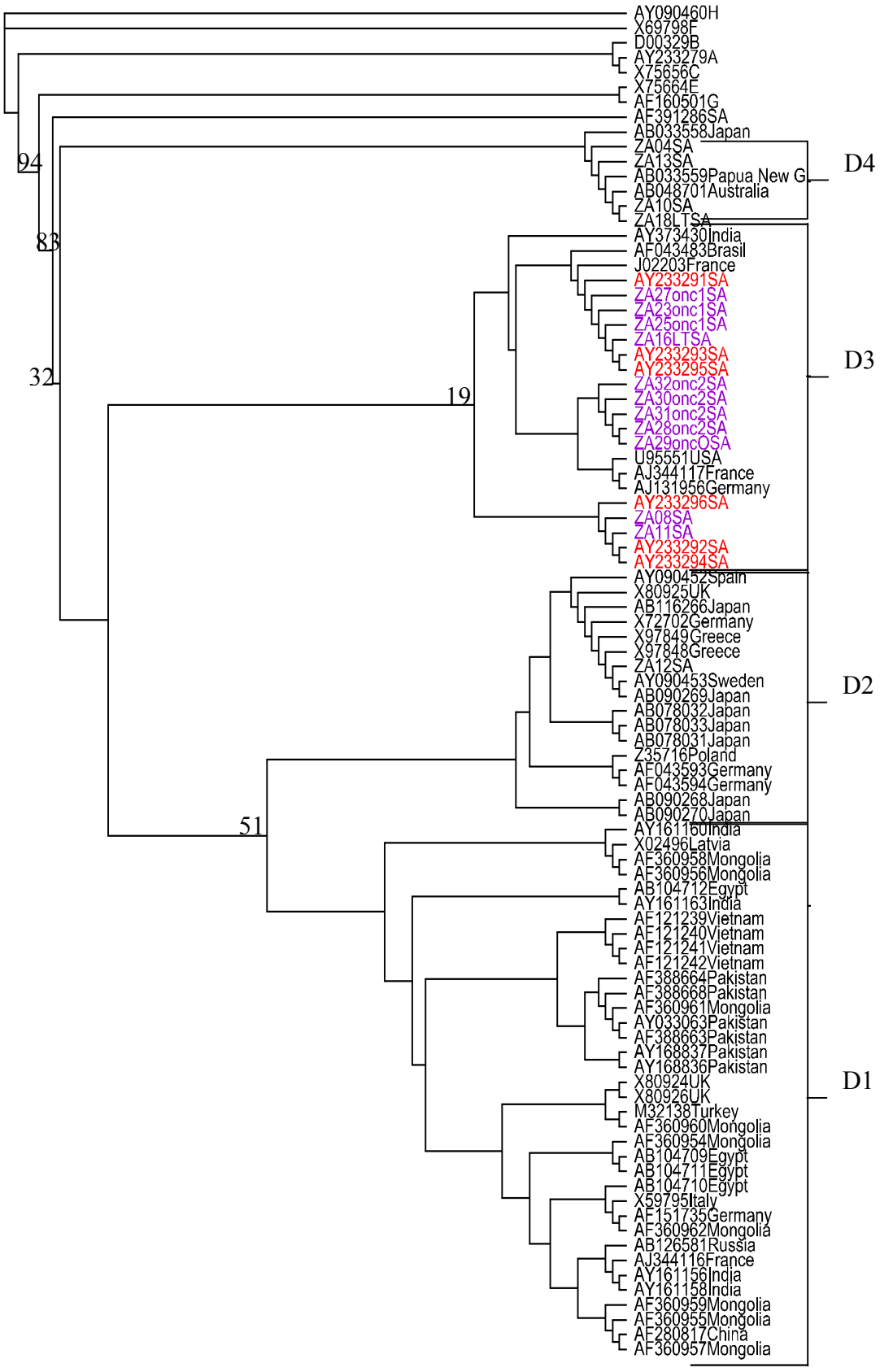

AY090460H
X69798F
D00329B
AY233279A
X75656C
X75664E
AE160501G
AF391286SA
AB033558Japan
ZA04SA
ZA13SA
AB033559Papua New G
AB048701Australia
ZA10SA
ZA18LTSA
AY373430India
AF043483Brasil
J02203France
AY233291SA
ZA27onc1SA
ZA23onc1SA
ZA25onc1SA
ZA16LTSA
AY233293SA
AY233295SA
ZA32onc2SA
ZA30onc2SA
ZA31onc2SA
ZA28onc2SA
ZA29oncQSA
U95551USA
AJ344117France
AJ131956Germany
AY233296SA
ZA08SA
ZA11SA
AY233292SA
AY233294SA
AY090452Spain
X80925UK
AB116266Japan
X72702Germany
X97849Greece
X97848Greece
ZA12SA
AY090453Sweden
AB090269Japan
AB078032Japan
AB078033Japan
AB078031Japan
Z35716Poland
AF043593Germany
AF043594Germany
AB090268Japan
AB090270Japan
AY161160India
X02496Latvia
AF360958Mongolia
AF360956Mongolia
AB104712Egypt
AY161163India
AF121239Vietnam
AF121240Vietnam
AF121241Vietnam
AF121242Vietnam
AF388664Pakistan
AF388668Pakistan
AF360961Mongolia
AY033063Pakistan
AF388663Pakistan
AY168837Pakistan
AY168836Pakistan
X80924UK
X80926UK
M32138Turkey
AF360960Mongolia
AF360954Mongolia
AB104709Egypt
AB104711Egypt
AB104710Egypt
X59795Italy
AF151735Germany
AF360962Mongolia
AB126581Russia
AJ344116France
AY161156India
AY161158India
AF360959Mongolia
AF360955Mongolia
AF280817China
AF360957Mongolia
94
83
32
19
51
D4
D3
D2
D1

Figure 4.8

4.2. DISCUSSION

This study provides the first complete genome sequences of genotype D isolates from South Africa. All Dsa isolates belong to a separate minor clade when phylogenetic analysis was performed at the level of the complete genome, polymerase and complete S genes. The intergroup divergence between Dsa and the remainder genotype D isolates was less than 4% at the level of the complete genome. Hence, Dsa could not be strictly classified as a subgenotype if we take the cut-off point defined by Bowyer *et al* (Bowyer *et al*., 1997). This was also true for D1, D2, and D3. Only D4 could be separated into a subgenotype, as the intergroup divergence of D4 with the remainding genotype D members was greater 4%.

In a later study analysis of complete genotype D genomes has classified genotype D isolates into 4 subgenotypes D1-D4, based on bootstrap values of >74% on a phylogenetic tree. This separation was not supported by significant bootstrap values at the level of the S gene (Norder *et al*., 2004). This was also the case in the present study (figure 4.3). According to this classification South African isolates clustered in subgenotype D3 and specified *ayw2*.

When phylogenetic analysis was carried out on the pre-S1 region of Dsa isolates and other isolates from GenBank, Dsa isolates clustered with most of the South African isolates from a previous study (Hardie and Williamson, 1997) (figure 4.7). A pattern of geographic distribution was also evident in the phylogenetic tree. Subgenotype D1 isolates were mostly from Asia, subgenotype D2 from Europe and Japan, subgenotype

D3 from South Africa and France, and subgenotype D4 from Oceania. South African isolates in clustering in D4 could have originated from Oceania as isolates from the study by Hardie and Williamson were obtained from individuals living in Cape Town. Cape Town is home to many immigrants from other regions of the World. The French Huguenots upon their arrival in South Africa in 1686 and 1726 could have introduced subgenotype D3 (http://www.orange-street-church.org/text/huguenot.htm). The clustering together of isolates from the same geographical region indicates low genetic variability. This is expected as the pre-S1 domain participates in binding of the viral particle to the hepatocyte during infection (Seeger and Mason, 2000) and therefore the human population group.

They were no amino acid changes in the pre-S1 that were unique to D3 isolates. The only amino acid change in the pre-S2 was Leu^{42} and this was specific to D3. Leu^{42} occurred within the amphipatic alpha helix between amino acids 41 and 52 of pre-S2, a region important in mediating cell permeability of pre-S2. This region may also play a part in internalization of the viral particle during infection (Oess and Hildt, 2000). Lys^{120} and Ile^{626} occurred within the binding region of HSP90 on the polymerase terminal protein and RT domain, respectively (Cho *et al*., 2000). Binding of HSP90 to HBV polymerase stabilizes the viral polymerase interaction with the encapsidation signal. This interaction is required early in replication for viral assembly and initiation of DNA synthesis through a protein-priming mechanism (Hu and Seeger, 1996). Leu^{343} of the spacer and Val^{463} of the RT occur within the fingers and palm of the DNA binding cleft of the HBV polymerase. These changes did not lead to any changes in

charge of the amino acid. Hence these changes may not translate to any differences in the DNA polymerase activity. None of the amino acids changes in the X region occurred within functionally important regions (Arii *et al*., 1992).

As in the case of the analysis of subgenotype A1 isolates there were many changes within *cis*-acting regulatory elements, but it was not possible to deduce the effect of these changes in the present study and would require further investigation.

4.3. CONCLUSION

In conclusion, analysis of the six isolates of genotype D from the present study and other genotype D isolates from GenBank has led to classification of genotype D isolates into four subgenotypes (D1- D4). The six isolates from the present study (Dsa) clustered together in a minor clade within subgenotype D3.

This analysis also led to the identification of differences at the level of the complete genome between genotype D isolates from South Africa and those from other regions of the world. These changes might affect replication as well as alter the outcome of disease in patients from South Africa. The intragroup divergence within genotype D isolates from South Africa was very low suggesting a recent introduction of Dsa as compared to subgenotype A1.

CHAPTER 5

INTEGRATION OF HBV SUBGENOTYPE A1 DNA INTO CHROMOSOMAL DNA DURING ACUTE HEPATITIS

SUMMARY

The integration of hepatitis B virus (HBV) DNA into chromosomal DNA in the liver of chronic HBV carriers has been widely documented. However, there is little evidence for integration during acute infection. This is largely the result of the ethical considerations with respect to performing liver biopsies on patients with acute hepatitis B. The aim of this study was to examine serum from black African patients with acute hepatitis B to determine if integrants of viral DNA can be detected in fragments of cellular DNA leaking from damaged hepatocytes into circulation.

DNA was extracted from serum of five patients with uncomplicated acute hepatitis B infection and one with fulminant hepatitis. Two subgenomic PCRs designed to amplify the complete genome of HBV were used and the resulting amplicons were cloned and sequenced.

Fortuitously, in addition to amplifying the subgenomic portions of the HBV viral genome, chromosomal DNA was also amplified from the serum of three patients with acute uncomplicated hepatitis B and one with fulminant hepatitis B. HBV DNA was found to be integrated into chromosome 7 q 11.23 in the WBSCR 1 gene in one of the acute HBV infected patients. The viral DNA portion was made up of 200 nucleotides covering the S and X genes in opposite orientation, with a nucleotide deletion of 1169

nucleotides. The identified viral host junction was situated at nucleotide 1774 in the cohesive overlap region of the viral genome, at a preferred topoisomerase 1 cleavage site. The chromosomal sequence was not rearranged. The patient made a full recovery and seroconverted to surface antibody-positivity. HBV DNA could not be amplified from his serum at this point.

Available Data suggest that integration of viral DNA into chromosomal DNA may occur rarely during acute hepatitis B. Clonal propagation of an integrant occurring during acute infection might play a role in the development of HCC.

5.1. INTRODUCTION

Epidemiological studies have shown that chronic infection with HBV is a major risk for the development of HCC (Beasley *et al*., 1981). Direct and indirect carcinogenic mechanisms have been implicated. Indirect carcinogenic mechanisms may be as a result of development of HCC from liver cirrhosis caused by HBV infection. Evidence for a direct effect is provided by the development of HCC in an otherwise normal liver (Paterson *et al*., 1985) and by the same observations in a variety on animal models. The latter includes association of HCC with acute and chronic HBV infection without liver cirrhosis in animals infected with other members of the family *Hepadnaviridae* (Marion *et al*., 1984, Popper *et al*., 1981) and in transgenic mice with the HBx gene of HBV incorporated into their germline (Kim *et al*., 1991).

It has been postulated that viral DNA integration may play a part in hepatocellular carcinogenesis. This is mainly because HBV integration into the host genome has been observed in 80-90% of HBV infected HCC patients and animals (Buendia, 1992, Hansen *et al*., 1993, Robinson, 1994). The details of how HBV integration causes carcinogenic changes are not well understood. One of the possible mechanisms is that the integration site may be within or close to a tumour related gene and clonal expansion may subsequently change the expression of the gene and cause HCC (Chami *et al*., 2000, Gozuacik *et al*., 2001, Wang *et al*., 1990).

If clonal expansion of integrated HBV DNA proves to be an important step in hepatocarcinogenesis, the time of integration of the viral genome into the host DNA becomes important. Because HCC typically develops in patients chronically infected with HBV, it has been postulated that integration occurs at some point during persistent infection. Further support comes from finding of integrated DNA in more than 80% of HBV related chronic hepatitis patients (Matsubara and Tokino, 1990).

In one of the studies in which HBV DNA integration during acute hepatitis was investigated Southern hybridisation and not sequencing was used to identify the integrants. The patient liver DNA was isolated and subjected to digestion with restriction enzymes, digested fragments were then separated on an agarose gel and transferred onto a membrane for Southern hybridization with a full length ^{32}P labelled HBV probe. Detection of bands at a position higher than 3.2 kb was indicative of integrated HBV DNA. Non-specific binding to chromosomal regions slightly homologous to viral genome could not be ruled out. Secondly serum of one of the acute patients tested positive for anti-HBc and anti-HBs only. This indicates that this was not the first time the patient was exposed to HBV infection; instead the patient might have been chronically infected, with a silent HBsAg state (Owiredu *et al*., 2001a). The remaining patient suffered from fulminant acute hepatitis.

In another study in which HBV DNA integration was investigated during acute hepatitis 3 out of 9 liver specimens from patients with acute hepatitis were established to contain viral DNA integrated into chromosomal DNA (Murakami *et al*., 2004). In 2

of the 3 patients integration occurred in a repetitive sequence. The viral host junction in three of the 4 integrants characterized from this patients was located in the cohesive overlap region between DR1 and DR2. This is a preferred site for integration of HBV viral DNA (Shih *et al*., 1987)

Attempts to demonstrate the integration of DHBV in duck during acute infection have been successful (Yang and Summers, 1999). Unlike in other hepadnaviruses, HCC does not, appear to be an outcome of DHBV infection (Buendia, 1992). The time at which integration occurred was studied in liver tissues and a similar study could not, for obvious ethical reasons, be performed in humans with acute hepatitis B infection. We have made use of the leakage of fragments of cellular DNA from damaged hepatocytes into plasma (Jahr *et al*., 2001) during liver injury to examine the serum of acute HBV infected patients for integrated DNA. HBV DNA present in serum could also come from damaged mononuclear cells (Murakami *et al*., 2004).

5.2. MATERIALS AND METHODS

5.2.1. PATIENTS STUDIED

Blood samples were collected from six South African Blacks suffering from acute hepatitis B. Five of the six patients (three males and two females) had uncomplicated illness, followed by complete recovery. The remaining one patient (female) had fulminant hepatitis B and died in liver failure. Serum alanine aminotransferase levels in these patients ranged between 799 IU/L and 4747 IU/L. The age range of the patients was 18-31 years. The sera were stored at -70ºC until analysed.

5.2.2. DNA EXTRACTION, AMPLIFICATION, CLONING AND SEQUENCING

Total DNA (genomic, mitochondrial, and viral) was extracted from the serum samples using the QIAamp DNA Mini Kit (QIAGEN GmbH., Hilden, Germany) according to the protocol provided by the manufacturer. A modification of the method described by Takahashi *et al* (Takahashi *et al*., 1998) used to amplify the complete viral genome. This involved the amplification of two overlapping fragments of the HBV genome 2.2 kb and 1.3 kb in length, respectively. The reaction mix for the amplification of the 2.2 kb fragment (fragment B) consisted of 2.25 µl of 10× Ex Taq buffer with 20mM $MgCl_2$, 2 µl of 2.5mM dNTP mix, 1.25 µl each of primers 1689(+) and 685(−), 2.5 µl of DNA, made up to 22.5 µl with water. The enzyme mix was made up of 1.875 µl of distilled water, 0.25 µl of 10× Ex Taq buffer, and 0.375 µl of TaKaRa Ex Taq polymerase. The 22.5 µl reaction mix was preheated to 94°C for 2 minutes and 2.25 µl TaKaRa Ex Taq enzyme mix was added at the first annealing step. This was followed by 40 cycles, cycling profile (table 2.1, page 62, chapter 2). For the amplification of the 1.3 kb

fragment (fragment A) using primers 1800(–) and 455(+) the annealing temperature was changed to 62°C for 30 seconds cycling profile table 2a. Both HBV-positive and HBV-negative controls were included. The former was a DNA extract from a known positive serum sample on the other hand the latter consisted of BQW (best quality water i.e injection water) water instead of DNA in the PCR mixture. To avoid cross-contamination and false-positive results, the precautions and procedures suggested by Kwok and Higuchi (Kwok and Higuchi, 1989) were strictly adhere to. DNA extraction, PCR amplification, and electrophoresis were performed in physically separated venues.

Amplicons were cloned into a pCR-Script^{TM} Amp SK^{+} vector (stratagene. La Jolla, CA, USA) according to the protocol provided by the manufacturer. The clones containing inserts were prepared for direct sequencing using the BigDye Terminator v3.0 Cycle Sequencing Ready Reaction Kit (Applied Biosystems., Foster City, USA) and sequenced on a 377 DNA automated sequencer (Applied Biosystems, Inc) using primers T3 (5'AATTAACCCTCACTAAAGGG3') AND T7 (5' GTAATACGACTCACTATAGGGC 3'). All sequences were analysed in both forward and reverse directions.

5.3. RESULTS

Analysis of the primers for specificity using the BLAST search (http://www.ncbi.nlm.nih.gov/BLAST/) showed that in addition to binding HBV DNA all the primers except 1687(+) could bind to chromosomal DNA (table 5.1).

Table 5.1. HBV "specific" primers and chromosomes complementary to these primers.

Primer	chromosome
455(+)	**4**
1800(-)	**1,2**,3,**7**,**11**,**12**,13,14,17,**20**,22,Y
685(-)	**1**,**2**,**4**,**7**,9,10,**11**,**12**,**20**,21,X

*Chromosomes shown in bold are complementary to both primers 1800(-) and 685(-). All chromosomes included have >60% of their bases complementary to a HBV primers.

In addition to amplifying the DNA fragments of the expected size, that is 1.3 kb and 2.2 kb for fragments A and B respectively, the two subgenomic polymerase chain reactions designed to amplify the complete genome of HBV, gave rise to smaller amplicons ranging in size from 0.1 kb to 0.7 kb (figure 5.1). These amplicons were successfully cloned into the pCR-Script^{TM} Amp SK^{+} vector figure and a number of clones from each of the six patients were sequenced (table 5.2).

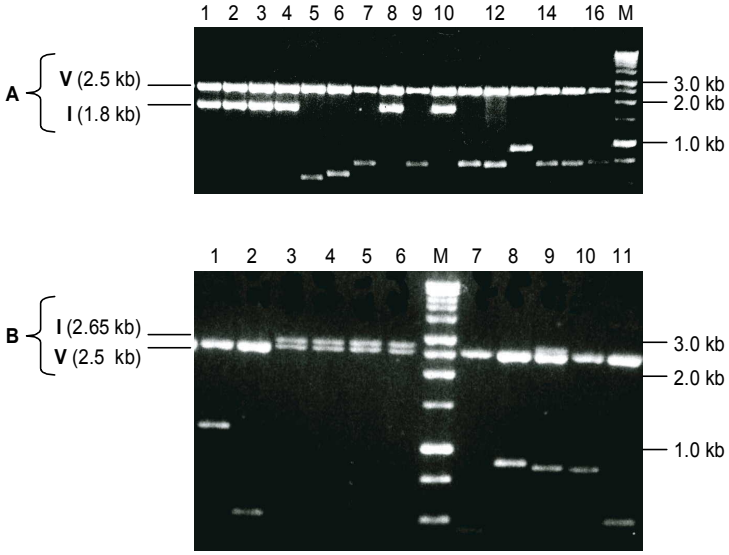


Figure 5.1. Ethidium bromide stained 1% agarose gel showing pPCR-Script Amp SK(+) plasmid containing amplicons restricted with *PvuII*. In panel A the fragments cloned were amplified with primers 455(+) and 1800(-). Lanes 1 to 4, 8 and 10 show the expected size insert of ~1.8 kb for fragment A amplicon (1.3 kb) whereas lanes 5 to 7, 9 and 11 to 16 show shorter inserts ranging in size 0.6 kb to 1.1 kb. In panel B the fragments cloned were amplified with primers 1687(+) and 685(-). Lanes 3 to 6 and 9 show the expected size insert of ~2.65 kb for fragment B amplicon (2.2 kb), lane 7 is vector alone and lanes 1,2, 8 to 11 show shorter inserts ranging in size 0.5 kb to 1.2 kb. M: Promega 1 kb molecular weight marker, V: vector, I: insert of expected size.

The majority of the clones contained HBV DNA only. These are analyzed in chapter 2 and 4. Fortuitously chromosomal DNA was amplified from all six patients with primer sets designed to amplify HBV DNA. In twelve cases chromosomal DNA was amplified

with primers designed to amplify HBV fragment A and in only one case the amplification of chromosomal DNA was a result of the PCR using primers for fragment B (table 5.2).

Table 5.2: Summary of cloning and sequencing results

Patient Details			**PCR fragment**[†]	**Number of clones**					**Chromosome number and length**
#	Sex	Age (y[‡])		Sequenced	With vector DNA	With HBV DNA only	With HBV/ chromosomal DNA	With Chromosomal DNA only	
355	F	18	B	5	1	3	0	1	7 (117bp[‡])
8225	M	23	A	3	0	2	0	1	1 (278 bp)
			B	4	0	4	0	0	-
28	M	26	A	4	0	2	0	2	4 (512 bp) 16 (140bp)
			B	4	0	4	0	0	-
4038	F	27	A	6	0	5	0	1	1 (97bp)
			B	4	0	4	0	0	-
5283[§]	F	32	A	12	2	6	0	4	4 (456 bp), 10 (235 bp), 10 (330 bp), 2 (422 bp)
			B	4	0	4	0	0	-
0962	M	27	A	8	0	4	1	3	1 (300 bp), 14 (410 bp), 17 (410 bp)
			B	5	0	5	0	0	-

[†] Fragment A amplified with primers 455 (+) and 1800 (-); fragment B amplified with primers 1687 (+) and 685 (-)

[‡] y – years; bp – base pairs, [§] Patient with fulminant hepatitis

HBV DNA was found integrated into host chromosome in one of the patients (figure 5.2). In this patient HBV DNA was found to be integrated into host chromosome 7q11.23 in the WBSCR 1 gene using primers for fragment. WBSCR1 corresponds to the eukaryotic initiation factor 4H identified in rabbits (Martindale *et al*., 2000, Richter-Cook *et al*., 1998) and is within the region encompassing the elastin gene. This region is commonly deleted in patients with Williams-Beuren syndrome, giving rise to loss of heterozygosity (Martindale *et al*., 2000, Osborne *et al*., 1996). The nucleotide sequence of the integrated HBV DNA fragment and the portion of the WBSCR1 gene sequenced have been deposited into GeneBank accession number AY233548. The length of the viral DNA was 200 nucleotides. This was made up of the S and the X genes in opposite orientation, and an 1169 nucleotide deletion between them (Figure 5.2). The right virus/host junction was situated at nucleotide 1774 in the cohesive overlap region of the HBV viral genome. This is a preferred topoisomerase 1 cleavage site. The left hand virus/host junction was not identified and therefore the length of the integrant could not be determined. The nucleotide sequence of the chromosomal DNA was not rearranged. Phylogenetic analysis of the one hundred and seventy five nucleotides of the S gene showed that the integrated viral DNA belonged to subgenotype A1 of genotype A which, is to be expected because subgenotype A1 is the predominant HBV genotype in South Africa (Kramvis *et al*., 2002). Translation of these 175 nucleotides in the correct frame (from the third nucleotide in the fragment) showed that the integrant was of serotype *ad*. The *w/r* determinant could not be determined because amino acid 160, which codes for this determinant, was not included in the fragment. Analysis of a complete HBV genome (GeneBank Accession No: AY233274) from the same patient

subsequently confirmed that the integrant belonged to subgenotype A1 and had an *adw2* serotype. The nucleotide sequence of the chromosomal DNA (262 bp) was completely conserved and not rearranged.

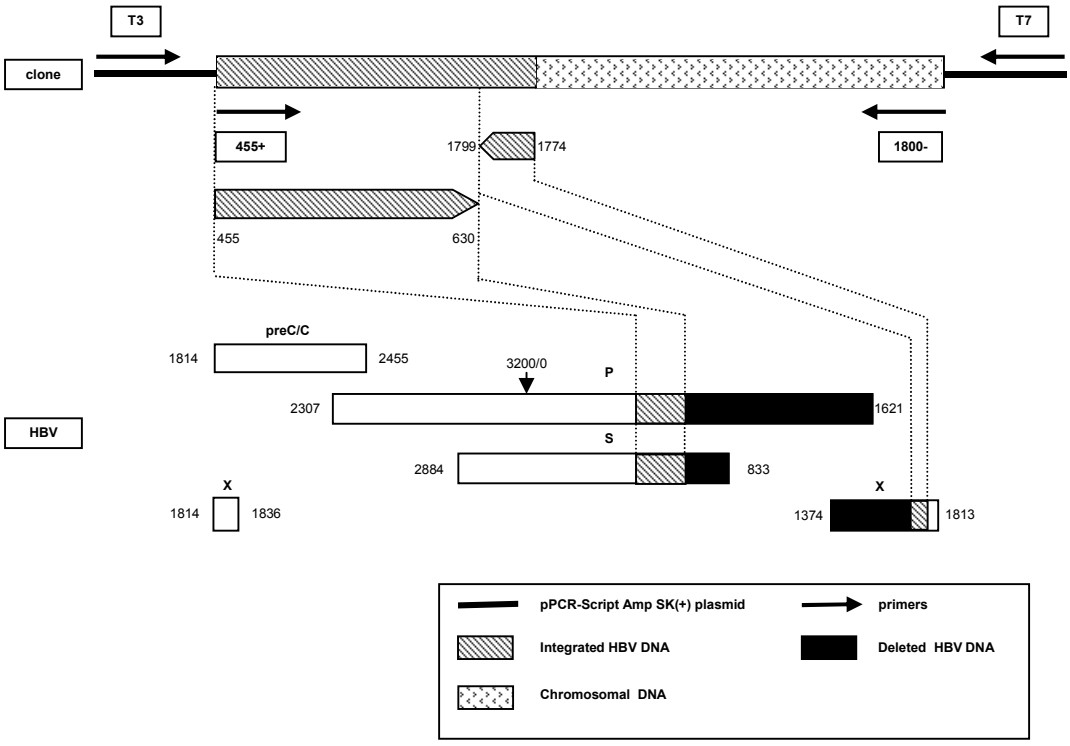


Figure 5.2. Schematic representation of the HBV DNA integrant amplified from the serum of acute hepatitis patient (#0962) using primers 455(+) and 1800(-) and cloned into pPCR-Script Amp SK(+) plasmid relative to the HBV genome. The lower part of the figure represents the genetic organization of the 4 open reading frames of HBV, preC/C: precore/core, P: polymerase gene, S: surface gene and X: X gene. Numbering according to nucleotide position of HBV GenBank accession # AY233276 where the *Eco*RI cleavage site is position 1.

The patient was a 27 year old Xhosa gold mine labourer, who presented with the typical features of uncomplicated acute viral hepatitis at the liver clinic of the Johannesburg general hospital. Table 5.3 below shows the biochemical liver function tests that were performed when he was first seen and the corresponding results.

Table 5.3. Liver function test performed on patient during acute hepatitis.

Liver function test	result
Serum bilirubin	257 µmol/L
Conjugated bilirubin	228 µmol/L
Alkaline phosphatase	91 U/L
γ-glutamyl transferase	94 U/L
Alanine amino transferase	3238 U/L
Aspartate aminotransferase	2781 U/L
Albumin	4 g/L

He also tested positive for HBsAg, HBeAg, and IgM core antibodies. After treatment the patient made a complete recovery with return of the biochemical test to normal and disappearance of HBsAg and HBeAg, accompanied by the appearance of IgG core antibodies, anti-HBs and anti-HBe. An attempt to amplify HBV DNA from the serum of the patient at this point was unsuccessful.

5.4. DISCUSSION

Although HCC development has been associated with HBV integration, the molecular mechanism of HBV integration and the role it plays in hepatocarcinogensis remain unclear. Unlike retroviruses, formation of a pro-virus after completion of the hepadnavirus replication cycle is not necessary (Seeger *et al*., 1986). However, some HBV DNA may enter the nucleus and can be integrated into the host genome through illegitimate recombination of linear viral DNA forms and host DNA, mediated by some cellular enzymes (Yang and Summers, 1999). Clonal expansion a of a cell containing integrated HBV DNA has been identified in a number of hepatocellular carcinomas that develop in the presence of chronic hepadnaviral infection (Kim *et al*., 1991, Marion *et al*., 1984, Popper *et al*., 1981). Unlike in woodchucks, no common chromosomal region has been associated with HBV integration (Hsu *et al*., 1988, Tokino and Matsubara, 1991).

Although we cannot exclude the possibility that the HBV integrant amplified was the result of a PCR artefact, there are a number of observations that argue against this interpretation and support our conviction that this integrant was authentic

1 The site of integration, in chromosome 7q11.23 is within the WBSCR1 gene and the integrant is within the region commonly deleted in patients with Williams-Beuren syndrome, giving rise to loss of heterozygosity (Martindale *et al*., 2000, Osborne *et al*., 1996). Additionally, the WBSCR 1 gene contains a high abundance of Alu repeats (Martindale *et al*., 2000), and the integration

occurred in one of these repetitive sequences. Repetitive elements have been shown to be preferred sites for integration of HBV DNA (Shaul *et al*., 1986, Murakami *et al*., 2004).

2 Analysis of the right hand virus/host junction revealed that integration occurred within the cohesive overlap region close to the 11 nucleotide direct repeat 1 (DR1) 1689-1700 on genotype. This is a preferred site for integration of HBV viral DNA (Shih *et al*., 1987).

3 The one viral DNA/host junction in the single integrant detected in our patient with acute hepatitis B was located at nucleotide 1774 of the viral genome, within a highly preferred topoisomerase I cleavage motif (TAA). The linear form hepadnaviral DNA is the primary substrate for integration (Wang and Rogler, 1991, Yang and Summers, 1999). Therefore the circular DNA must first be linearized and then imported into the nucleus of the infected cell before integration can occur (Gong *et al*., 1999, Gong *et al*., 1995, Wang and Rogler, 1991). Wang and Rogler (Wang and Rogler, 1991) have produced experimental evidence for a key role for the cellular enzyme topoisomerase I in the illegitimate recombination between hepadnaviral and host DNA. A role for this enzyme in illegitimate recombination is supported both by the finding of highly preferred topoisomerase I cleavage motifs in the immediate vicinity of almost all crossover sites in mammalian DNA, and by the observation that integrants in

WHV-induced HCC in woodchucks occurred at preferred topoisomerase I cleavage sites (Wang and Rogler, 1991).

4 The integrated HBV DNA covered the S and X regions, the open reading frames most commonly integrated into chromosomal DNA (Unsal *et al*., 1994). The X protein and the HBsAg can function as a transactivators (Colgrove et al., 1989, Twu and Schloemer, 1987, Wang *et al*., 2004, Zahm *et al*., 1988), this is strengthened by the fact that 90% of transgenic mice expressing HBx developed HCC (Kim *et al*., 1991).

5 Finally, the integrant detected in the serum of our patient with acute hepatitis B shared similar characteristics to the integrants described in HCC developing in the presence of chronic HBV infection (Slagle *et al*., 1992). It contained a long deletion (1169 nucleotides in length), and the viral DNA between the right end of the deletion and the right virus/chromosome junction was an inverted X sequence (figure 5.2). One or more deletions or other more complex rearrangements are invariably found in inserted hepadnaviral DNA (Ogston *et al*., 1982, Slagle *et al*., 1992), indicating that specific mechanisms for integration of functional viral DNA into chromosomal DNA are not encoded by the virus. Simple microdeletions are most common, probably because the single-strand ends of viral DNA resulting from topoisomerase I cleavage are sensitive to cellular nucleases. The larger deletions and complex rearrangements of the viral DNA are generally believed to occur after integration (Rogler and

Summers, 1982), although WHV with extensive rearrangements of DNA has been described in hepatocyte nuclei in woodchucks (Rogler and Summers, 1984) and there is some evidence in these animals that DNA rearranged in this way may be preferentially integrated (Kew *et al*., 1993). Similarly, Yaginuma *et al* (Yaginuma *et al*., 1987) observed features of HBV integration to be common to HCC and patients with chronic active hepatitis and an analogous integrant to the one detected in this study has been described in a 9 year old child with HCC (Tsuei *et al*., 1994).

Integration of hepadnaviral DNA into host DNA is normally rare, perhaps because recycling of viral DNA into the nucleus is at a low level during productive infection (Yang and Summers, 1999). The frequency of integration of DHBV in experimentally infected ducklings was estimated to be one viral genome per 10^3 to 10^4 cells by 6 days after infection (Yang and Summers, 1999). It is not known if these recombination events occurred in many cells or if a few cells produced many such events. In keeping with the belief that insertion events are uncommon, we found only a single integrant among many clones in only one of 6 patients with acute hepatitis B. Our finding, and that of others (Murakami *et al*., 2004, Yang and Summers, 1999), of integrated viral DNA in early infection is in keeping with the observation that importation of linear DNA into the nucleus is essential for insertion of viral DNA into chromosomal DNA, and that such importation is known to occur in the phase of cccDNA amplification during the initiation of infection (Tuttleman *et al*., 1986). Opportunities for viral integration beyond the early phase of infection would require continued importation of

viral DNA into the nucleus, and would depend on new rounds of infection or loss and replacement of cccDNA by cell turnover (Yang and Summers, 1999). Thus, insertion of linear forms of viral DNA may occur preferentially during early phases of infection or during periods of extensive hepatocyte turnover and/or re-infection, at times when linear viral DNA is imported into nuclei. Integration may be enhanced during these periods by the availability of DNA ends resulting from DNA replication or damage (Petersen *et al*., 1997). Alternatively, interference with the normal replication cycle at some stage in chronic infection may result in the accumulation of linear forms of DNA, and there is some evidence for defective replication of HBV and accumulation of replicative intermediates in patients with HBV-associated HCC (Raimondo *et al*., 1988).

The clinical, biochemical, and serological features of the acute hepatitis B in the patient with the integrant did not differ obviously from those in the other patients with uncomplicated disease. He made a complete clinical and biochemical recovery, and seroconverted to anti-HBs and IgG anti-HBc. At this stage HBV DNA could no longer be detected in his serum by PCR amplification and further cloning was not attempted. Even if the integrated HBV DNA was still present in hepatocytes, it would not be detected in peripheral blood once healing of hepatocyte cell membranes had taken place. Thus, persistence of the integrated HBV DNA could only have been detected by analysis of liver tissue, and a liver biopsy was not felt to be justified. Although sources other than the liver for the chromosomal DNA amplified in the peripheral blood of the six patients with acute hepatitis B (including the one with fulminant hepatitis) are

possible (for example, peripheral blood mononuclear cells), the finding of viral DNA integrated into host DNA in one of the patients favors leakage of chromosomal DNA fragments from damaged hepatocytes during the acute liver injury.

5.5. CONCLUSION

Our finding of an integrant in a patient with acute hepatitis B may have no significance in relation to subsequent HCC development. Clonal expansion of an integrant is required for tumor formation, and the finding of a viral insertion in a tumor may simply reflect the fact that the clonal progenitor of the tumor happened to carry a viral integration (Ogston *et al*., 1982). Moreover, integrants may be lost from successive generations of cells (Gong *et al*., 1996). Nevertheless, it is in keeping with the observation made in ducks that hepadnaviral DNA insertion may occur at a very early stage of infection, and raises the possibility that in certain circumstances early integration of HBV DNA might play a role in HBV-induced hepatocarcinogenesis.

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSION

The global distribution of chronic HBV infection, as well as HCC, is geographically distinct. This suggests that different genotypes may be associated with differences in pathogenesis of HBV infection. Black South Africans are predominantly infected with genotype A isolates of HBV, this is in contrary to other African populations south of the Sahara who are mostly infected with genotype E.

Complete genome amplification of HBV isolates allows for the amplification of the entire genome of a single HBV virus. This is particularly important, as the HBV usually exist as quasispecies. Amplification of the 23 complete HBV genomes in the present study will greatly improve the data available on complete HBV genomes from Africa, especially as subgenotype A1 may exist in other parts of Africa. The cloning of these complete genomes may serve as a backbone for construction of plasmids in the future for analysis of subgenotype A1 and genotype Dsa isolates *in vitro*.

The detailed analysis of fifteen subgenotype A1 isolates from South Africa in the present study has confirmed many previously reported changes peculiar to this subgenotype and has identified new ones.

The G1888A substitution was shown to occur only in subgenotype A1 isolates. *In vitro* transcription/translation experiments could not demonstrate the effect of the G1888A mutation on translation of the core protein. Use of a more sensitive flow cytometric

method led to the observation that the G1888A mutation lead to a 18.75% decrease in core protein translation. This could be as a result of the following:

1. A number of the scanning ribosomes may bind to the newly introduced 1888 AUG, hence reducing the amount of ribosomes reaching the down stream core AUG where the translation of the core protein is initiated.
2. The termination codon of the uORF, created by the 1888A mutation is very close to the core AUG. This can result in the accumulation of the terminating ribosomes at this stop codon, which may interfere with the initiation of translation at the core AUG, by a steric hindrance mechanism.

Characterization of a large number of subgenotype A1 isolates let to the identification of double or triple point mutations at position 1809-1812 of these isolates. These mutations exist as a stable trait in subgenotype A1 i.e they do not develop during the course of infection. The most common combination of the mutations the 1809T/1812T has been shown to result in a 20% decrease in HBeAg production (Ahn *et al*., 2003). HBeAg functions as a T cell tolerogen and regulates the immune response to the intracellular nucleocapsid. This is achieved by eliciting tolerance in HBcAg/HBeAg-specific T cells. The decrease in core protein translation (as a result of the G1888A substitution) immediately after infection might lead to fewer core antigen epitopes being recognized, leading to a less aggressive immune response. This may compensate for the decrease in production of the immuno-suppresive HBeAg (Chen *et al*., 2004) as a result of 1809T 1812T mutations resulting in a less severe response against the virus.

An inadequate immune response to the virus might result in incomplete clearance of the virus leading to chronic (persistent) infection.

$Gln^{334}$and Lys^{338} were unique to subgenotype A1. This aa substitutions occurred in the fingers of the HBV polymerase, within the positively charged DNA-binding cleft and resulted in a change in charge of this domain, possibly affecting the binding of the viral polymerase to DNA hence DNA synthesis. Sera from patients infected with subgenotype A1 have been shown to contain lower levels of DNA (Tanaka *et al*., 2004, Kramvis *et al*., 1997). Persistence of the virus as a result of an inadequate immune response as well as low viral DNA levels may ultimately result in development of HCC. This might explain the higher hepatocarcinogenic potential of subgenotype A1 in South Africa (Kew *et al*., 2005).

The six complete genotype D sequences from the present study represent the only complete genotype D genomes characterized from this region to date. These isolates clustered together in a separate clade (Dsa). Nucleotide changes and aa substitutions specific to Dsa were also identified in the *cis*-acting elements and different ORFs respectively.

The identification of integrated viral DNA in early infection is crucial to the understanding of HCC development. This is one of the few times that an HBV integrant has been described in the acute phase of the disease. This discovery raises the possibility that in certain circumstances early integration of HBV DNA into

chromosomal DNA might play a role in HBV-induced hepatocarcinogenesis (Kew *et al*., 2005).

Although HBV is hyperendemic in southern African Blacks and there is a correspondingly high incidence of hepatocellular carcinoma, prior to this study there was a paucity of data on the complete genomes of strains circulating in this population. Genotype A, specifically subgenotype A1 was found to predominate over genotype D. The sequence of subgenotype A1 and southern African genotype D isolates has been fully characterized. The higher intragroup divergence of subgenotype A1 would suggest that this strain is endemic for a longer time in the southern African population whereas the lower intragroup divergence of genotype D isolates point to a more recent introduction of these strains into the population. Work carried out in the Molecular Hepatology Research Unit, University of the Witwatersrand, has shown that subgenotype A1 is more hepatocarcinogenic than subgenotype A2 and other genotypes found in southern Africa. The variations and mutations in subgenotype A1, which may result in the low replication rate, high HBeAg-negativity and persistence of this strain, were confirmed and identified in this study and may contribute to the high hepatocarcinogenic potential of this strain. Fortuitously, HBV belonging to subgenotype A1 has been found integrated in the WBSCR1 gene of chromosome 7 in the serum of an acute hepatitis B patient, intimating that integration can take place during the early stages of infection. These findings are significant in our quest to understand, at a molecular level, the mechanisms involved in the pathogenesis of HCC in southern African blacks, in relation to the HBV.

APPENDICES

A SOLUTIONS, BUFFERS, AND GELS.

A1.1. *1% Agarose*

1 g agarose
10 mls 10x TBE
Make up to 100 ml with distilled water

A1.2. *10x Tris Boric Acid EDTA Buffer (TBE)*

2.5 litres of 10x TBE
270 g Tris base
127.5 g Boric acid
18.6 g of EDTA
Make up to 2.5 litres with distilled water
Autoclave

A1.3. *1 litre of 50xTAE*

242 g Tris base
57.1 ml glacial acetic acid
37.2 g of Na_2EDTA
Make up to 1 litre with distilled water

A1.4. *6x DNA Loading Buffer*

3 mls glycerol,
25 mg Bromophenol (Merk, Germany)
Make up to 10 ml with distilled water.

A1.5. *30% acrylamide solution*

30 g acrylamide (Sigma, Germany)
0.8 g bisacrylamide (Promega, USA)
Make up to 100 ml with distilled water

A1.6. *15% Protein gel mix*

96 mls water
100 mls 1.5M Tris pH 8.8
2 mls 20% SDS
For 1 gel use 4.9ml of this plus
5 ml of 30% acrylamide (Sigma, Germany)
0.1 ml of 10% APS (Ammonium persulphate, Promega, USA)
4 µl of TEMED (Promega, USA)

A1.7. *Stacking gel mix*

139.75mls distilled water
63 ml 0.5M Tris pH 6.8
1.25 ml 20% SDS
For 1 gel use 2.04 ml of t his plus
415 µl of 30% acrylamide (Sigma, Germany)
25 µl of 10% APS (Ammonium persulphate, Promega, USA)
2.5 µl of TEMED (Promega, USA)

A1.8. *Protein running buffer 10x*

75 g Tris base
470 g Glycine
25 g SDS
Make up to 2.5 litres with distilled water

A1.9. *Protein loading buffer 5x*

50 ml 1M Tris pH 6.8
30 g SDS
1g bromophenol blue
100ml glycerol
Make up to 250 ml with distilled water
Add 100 µl of 2M dithiothreitol (DTT) to 400 ml.

A1.10. *Fixing buffer*

50% Methanol
10% Glacial acetic acid
40% distilled water

A1.11. *1X EDTA*

1 g EDTA (Saarchem, RSA)
10 ml PBS
Filter sterilize

B BACTERIAL CULTURE MEDIA

B1.1. *Ampicillin (50 mg/ml)*

1 g ampicillin
Make up to a final volume of 20 ml with distilled water
Sterilize by filtration
Store in 1 ml aliquots at -20^{o}C

B1.2. *5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal) (20mg/ml)*

100mg X-Gal (Sigma, Germany)
5 ml of dimethyl formamide
Filter sterilize
Store ate -20 ^{o}C

B1.3. *isopropyl-beta-D-thiogalactopyranoside (IPTG) (200mg/ml)*

1 g IPTG (Sigma, Germany)
5 ml of distilled water
Store at -20 ^{o}C

B1.4. *Luria-Bertini medium (LB)(1 litre)*

10 g Tryptone (Pronadisa, Madrid)
5g yeast extract (Pronadisa, Madrid)
10 g NaCl
Make up to 1 litre with distilled water
Autoclace

B1.5. *Luria-Bertini medium agar plates with ampicillin*

15 g agar (Pronadisa, Madrid)
10 g Tryptone (Pronadisa, Madrid)
5 g yeast extract (Pronadisa, Madrid)
10 g NaCl
Make up to 1 litre with distilled water
Autoclace
Cool to 55 ^{o}C add 1 mls of ampicillin (50mg/ml)
Pour into 60mm petri dishes and store at 4 ^{o}C

B1.6. *Luria-Bertini medium agar plates with X-gal*

Spread 40 µl of X-gal followed by 4 µl of IPTG on LB agar plates.

C PREPARATION OF CELL CULTURE MEDIUM

C1.1. 10.4 g Powdered RPMI 1640 + L-Glutamine (Gibco Invitrogen, UK)
1.19 g Hepes
2 g $NaHCO_3$
Dissolve in 1 ℓ of distilled water } pH 7.12

Add the following filter sterilized supplements;

1 ml Na_2SeO_3 (3 x 10^{-5} M)
1 ml $FeSO_4.7H_2O$ (1 x 10^{-4} M)
1 ml of solution 3 which contains:$(NH_4)_6MO_7O_2.4H_2O$ (3 x 10^{-6} M)/ $MnCl_2.4H_2O$ (3 x 10^{-7} M)/ NH_4VO_3 (1 x 10^{-5} M)
100 µl Linoleic/Oleic acid BSA complex (3 x 10^{-5} M)
1 ml Ethanolamine (3 x 10^{-2} M)
Sterilize by membrane filtration using 0.22µm filters.
Store at -20 ^{o}C

C1.2. Preparation of different stock solution.

C1.2.1. Na_2SeO_3 (3 x 10^{-5} M)
Dissolved 518.7mg of Na_2SeO_3 (Sigma, Germany) in 100 ml distilled water and dilute at 1/1000.

C1.2.2. $FeSO_4.7H_2O$ (1 x 10^{-4} M)

Dissolve 278mg $FeSO_4.7H_2O$ (Sigma, Germany) in 100 ml distilled water, add 1 drop of Concentrated HCL, and then dilute at 1/100.

C1.2.3. $(NH_4)_6MO_7O_2.4H_2O$ (3 x 10^{-6} M)

Dissolve 37.1mg :$(NH_4)_6MO_7O_2.4H_2O$ (Sigma, Germany) in 100 ml of distilled water, dissolve 59.4 mg of $MnCl_2.4H_2O$ (Sigman, Germany) in 100 ml of distilled water and dilute to 1/100, dissolve 11.7 mg NH_4VO_3 (Sigma, Germany) in 100 ml of distilled water. Combine the three solutions together.

C1.2.4. Linoleic/Oleic acid BSA complex (3 x 10^{-5} M)

Dissolve 302µg linoleic acid and 10mg linoleic acid-water solble (Sigma, Germany), 305.8µg oleic acid and 9.38mg oleic acid solble (Sigma, Germany) in 36 ml of distilled water.

C1.2.5. Ethanolamine (3 x 10^{-2} M)

Dissolve 183.2 mg ethanolamine (Sigma, Germany) in 100 ml of distilled water.

Table 7.1. Flow cytometry analysis of wild-type GFP and uORF + GFP

Plasmid	GFP counts in 1st set of experiments	GFP counts in 2nd set of experiments	GFP counts in 3rd set of experiments
wild-type GFP	16.2, 17.0, 15.7	27.4, 28.5, 29.8	27.7, 25.3, 25.5
uORF + GFP	14.3, 15.5, 13.7	24.4, 24.6, 23.2	23.4, 24.8, 22.5

Table 7.2. Flow cytometry analysis of core/GFP fusion plasmids with or without the G1888A mutation.

Plasmid	GFP counts in 1st set of experiments	GFP counts in 2nd set of experiments	GFP counts in 3rd set of experiments
CoreA1	7.9, 9.2, 8.9	7.7, 7.0, 7.6	9.5, 8.4, 7.4
CoreA2	7.2, 7.0, 7.4	6.8, 6.3, 6.4	6.2, 6.2, 6.4
Core A1+Opt	3.6, 3.6, 3.5	3.6, 3.3, 3.6	3.7, 3.5, 3.6

Table 7.3. Flow cytometry analysis of core/GFP-expressing plasmids with G1888A substitution and varying distances of the uORF stop codon from the core AUG.

Plasmid	GFP counts in 1st set of experiments	GFP counts in 2nd set of experiments	GFP counts in 3rd set of experiments
G1888A+Opt	3.1, 3.6, 3.7	3.5, 3.5,3.5	3.6, 3.8, 3.8
G1888A(15)	5.1, 4.9, 5.1	4.9, 4.9, 5.0	4.9, 4.9, 4.84
G1888A(60)	4.7, 4.7, 4.8	4.9, 4.9, 5.0	5.2, 5.1, 5.2
G1888A(100)	5.4, 5.0, 5.1	5.2, 5.1, 5.2	5.2, 5.4, 5.3