Hepatitis B Virus:

A Longitudinal Study

A thesis presented in partial fulfilment of the requirements
for the degree of

Masters of Science

At

Massey University,
Palmerston North, New Zealand

Katherine Graciosa Blake-Palmer

2002

Abstract

Hepatitis B Virus (HBV) is a member of the hepadnavirus family. Viruses from this family infect primate, rodent and avian species. Wild type HBV virions consist of partially double-stranded circular DNA which is converted into covalently closed circular molecules in nuclei upon infection into host cells. The HBV genome is about 3.2kb in size and consists of four transcripts encoding the surface, core, polymerase and X proteins, in overlapping reading frames. HBV infection causes a variety of liver diseases in humans, for example, liver cirrhosis and hepatocellular carcinoma. Clinical manifestations range from asymptomatic to acute. The outcome of acute hepatitis B infection may be influenced by host factors some of which are controlled by the Major Histocompatibility Complex (MHC). In humans the MHC is known as the Human Leukocyte Antigen (HLA) region. Accordingly, the individuals involved in this study were HLA typed.

The aim of this study is to investigate HBV DNA differences in three different clinical types of hepatitis B disease over a 15 year period, and to determine if there is a correlation between specific HBV variants and particular clinical states. In 1985, 93% of the population of Kawerau (7,901) was tested for HBV, those found to be positive (519) have been monitored ever since. In 1998, individuals that fitted our requirements were invited to participate in our study. HBV DNA was extracted from blood samples and complete genomes sequenced, over 120,000 nucleotides were sequenced. Differences in HBV genomes sequenced between clinical types and HBV genotypes were compared. HLA alleles between the different clinical types were compared, as well as comparing HBV infected individuals with the general New Zealand population. The overall project is a major one and the results of this thesis get it well underway.

Acknowledgements

I would like to thank Dr Chris Moyes of the Whakatane Child Health and Hepatitis Foundation for providing the hepatitis B virus samples.

Thank you David Penny for your supervision and much-needed assistance in writing this thesis.

Abby Harrison – thanks so much for your continued support throughout these last two years!

Thanks also to Trish McLenachan for all your help in the lab.

Thanks also to Rissa Otta for your help with the statistical analysis.

A big thanks to all my family and friends for your support outside of the lab – everybody needs to escape now and then!

Contents

Chapter 3 Materials and Methods

Chapter 4 Results and Discussion

List of Figures

List of Tables

Appendix A

Appendix B

Chapter 1 Introduction

A unique opportunity to study the hepatitis B virus (HBV) arises from having blood samples taken in 1985 and 1998 from the same human hosts. This thesis explores the changes that occur in the hepatitis B virus over that 15 year period. Viruses from hosts belonging to three clinical groups were sequenced and analysed together with the immune genes of the human host. This was to see if a particular immune type responds better to hepatitis B infection than others. This introduction covers two parts; part one is a general introduction to the virus, and part two is an introduction to the Human Leukocyte Antigen (HLA) complex.

1.1 General introduction to HBV

Hepatitis caused by the hepatitis B virus (HBV) is a major disease world wide, there are estimated to be about 300 million carriers, the majority of which are in South East Asia and adjacent areas including Oceania. HBV infection is highly variable due to complex interactions between the host and the virus itself. HBV is a coated, partially double-stranded DNA virus, which is remarkable in that it replicates through an RNA intermediate with reverse transcriptase. The HBV genome consists of approximately 3200 base pairs and contains a number of transcripts that are described below (figure 1.1).

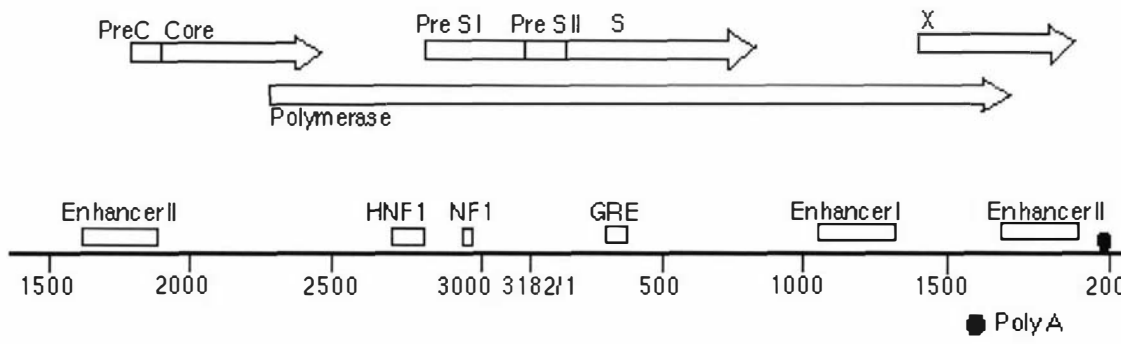

Figure 1.1 HBV genome. HNFI = HNFI binding site; NF1 = NF1 binding site; GRE = Glucocorticoid receptor binding site. Nucleotide numbering according to (Galibert et al., 1979). Source: modified from (Schaller and Fischer, 1991)

Two types of transcriptional products result from viral replication, mRNAs for protein synthesis, and the pregenome RNA for viral genome replication. For mRNA there are four transcripts that encode the following viral proteins (Table 1.1, and Figure 1.1).

- The Precore/Core gene transcript

The precore/core gene transcript codes for two viral proteins. One makes up the nucleocapsid and containing the hepatitis B core antigen (HBcAg). The other protein becomes (after post-translational processing) the hepatitis B e-antigen (HBeAg). The HBeAg function is not yet understood, but because its presence is measurable in blood, it is used as a marker for ongoing viral replication and infectivity.

- The surface genes

The surface gene transcript codes for 3 lipoproteins (Large, Medium and Small) which make up the viral envelope. All three proteins contain the hepatitis B surface antigen (HBsAg) (Figure 1.1).

- The Polymerase gene transcript

The polymerase gene transcript codes for the viral DNA polymerase, a reverse transcriptase and Ribonuclease H that are essential for viral reproduction.

- X gene transcript

The X gene has several regulatory functions for the virus, and transactivator effects.

Each gene region is discussed in more detail in Chapter 2.

1.1.1 Virus lifecycle

An overview of the virus lifecycle is shown in figure 1.2. HBV preferentially enters and survives in hepatocytes. The viral structures responsible for the attachment of HBV to the host cell surface are not yet fully characterised; it is believed that they are most likely

located in the pre SI and pre SII proteins (Grob, 1998). Various candidates for the receptor structure on hepatocytes enabling viral entry have been postulated, a recent one being Liver Annexin VI (Grob, 1998).

The infectious virus is a 42 nm virion. The virion contains within it the nucleocapsid and within the nucleocapsid is a full length minus strand of DNA and a partial plus strand of DNA (of various lengths) together with enzymes such as the viral DNA polymerase. The nucleocapsid enters the cell, leaving the S antigen encrusted virus-envelope attached to the host cell surface. The nucleocapsid enters the host cell nucleus, where the endogenous viral polymerase completes the plus strand, forming a covalently closed circular (ccc) viral genome of approximately 3.2 kilo-bases in size.

Table 1.1 Hepatitis B transcripts and their products.

Transcript	Size (nt)	Product	Mapped Bases
Pregenome	3500	Viral DNA pregenome	1903-1-1902
Precore, Core mRNA		Precore protein (HBeAg)	1816-2454
		Core protein (HBcAg)	1903-2454
		Polymerase	2309-1-1625
Pre S I mRNA	2400	Large protein	2850-1-837
Pre S II mRNA	2100	Medium and	3207-1-837,
S mRNA	680	Small proteins	157-837
X mRNA	650	X protein	1376-1840

The transcripts are produced by the nuclear transcription machinery of the host cell, and all terminate at a common site (Gunther et al., 1999). The transcripts are transported to the cytoplasm where they are translated. Once in protein form they are either incorporated into an infectious virion, or in the case of precore, core and surface proteins, can be secreted via the endoplasmic reticulum as non-infectious particles (Mason and Seeger, 1991).

For viral replication, the full genome is copied into an RNA pregenome. This is then transported into the cytoplasm, where the core particle, pregenomic RNA and the viral polymerase self-assemble into the nucleocapsid. Within the nucleocapsid the reverse transcriptase converts the pregenomic RNA into a minus strand of DNA and a partial plus

strand of DNA. The nucleocapsid then migrates to the endoplasmic reticulum, where it is enveloped in surface envelope proteins. The now infectious virion is released from the cell through vesicle-mediated transport (Pugh and Bassendine, 1990). Alternatively instead of being released outside the cell, the nucleocapsid particles can disassemble and re-release nascent viral genomes in the original host cell. These genomes can re-enter the nucleus and mature to covalently closed circular DNA molecules, amplifying the pool of transcriptional templates (Pugh and Bassendine, 1990).

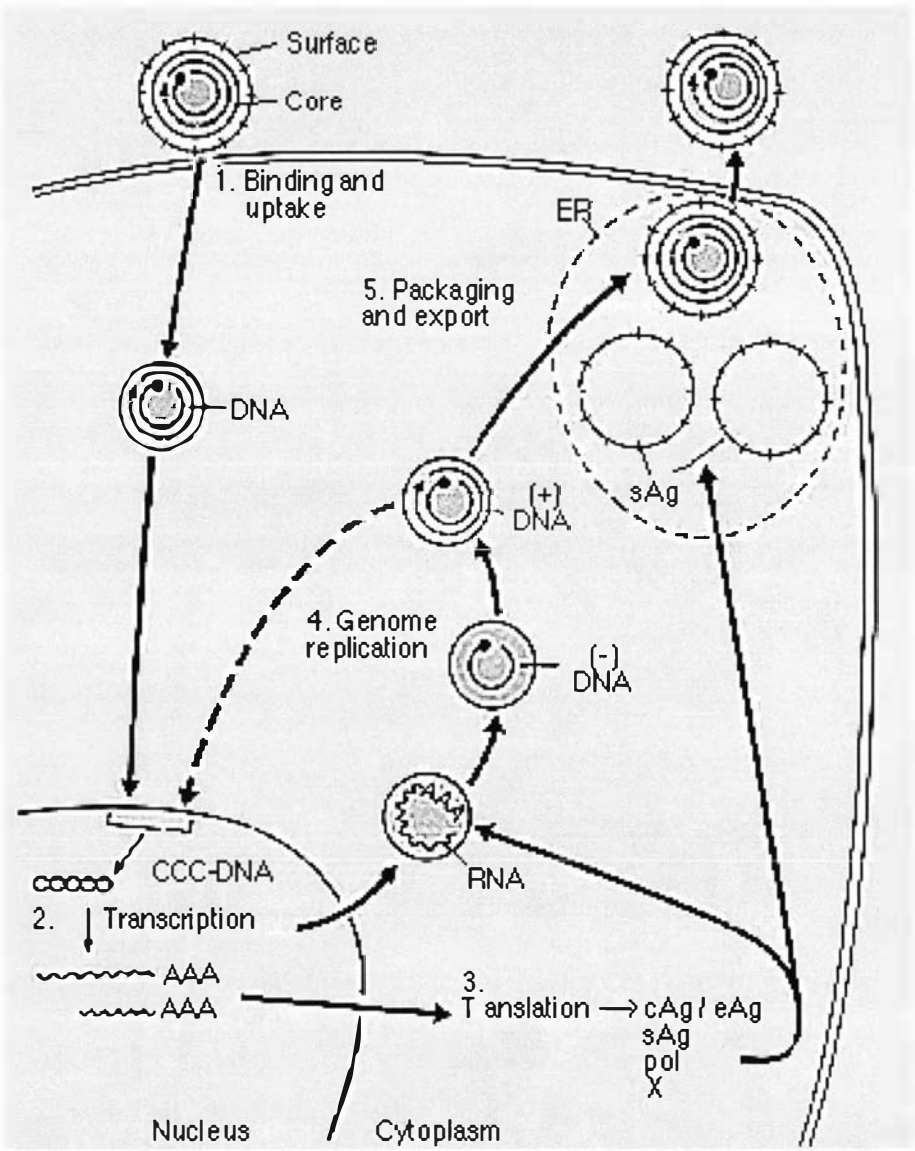

Figure 1.2 Lifecycle of the virus. 1. Virus is taken up by the cell. 2. and 3. Transcription of DNA produces mRNAs which are translated to protein products. 4. The pregenome RNA is packaged into core structures plus viral enzymes (e.g. Polymerase). 5. Mature

cores are packaged into surface antigen particles, which accumulate in the endoplasmic reticulum (ER), and are exported from the cell. Source (Pugh and Bassendine, 1990).

1.1.2 Viral infection

HBV is transmitted through body fluids, such as blood and sexual fluids. It is an exceptionally hardy virus, indeed it is more infectious than HIV. Consequently it is easily spread through contaminated objects such as dirty needles. There are two types of transmission, vertical, where the virus is passed on from mother through to infant, if this occurs at birth or shortly after it is known as perinatal transfer. All other transmissions are horizontal.

Virus infection may result in a broad spectrum of liver diseases, ranging from an asymptomatic carrier state with no liver dysfunction to fatal fulminant hepatitis. Adult infection usually leads to acute liver disease. Acute hepatitis B is defined as a transient liver disease which can resolve spontaneously (acute self-limited hepatitis) (Gunther et al., 1999). After an acute infection of HBV, over 90% of adults will appear to recover, and the HBV surface antigen (HBsAg) will be cleared from the serum. HBV clearance is associated with both a $CD8^+$ cytotoxic T lymphocyte (CTL) response directed against the infected liver cell and formation of antibodies to the HBsAg (Thio et al., 1999). The 10% of adults who do not recover from HBV infection become chronic carriers, and are predisposed to hepatic failure and heptocellular carcinoma. Fulminant hepatitis is fortunately rare, as it is associated with high mortality and is often only treatable by liver transplantation (Gunther et al., 1999).

Chronic HBV infection usually results in the presence of both HBsAg and HBeAg in serum, high viral load, absent or minimal inflammatory liver disease, and sometimes even a lack of antibody response to HBcAg. Ninety percent of perinatal transfers result in chronic infection (Gunther et al., 1999). Chronic infection is thought to occur due to a low or absent immune response against HBV (Gunther et al., 1999), otherwise known as the "immune tolerant" phase. This immune tolerant phase can last up to several decades

until an inflammatory liver disease develops, which is frequently followed by seroconversion from HBeAg positive to HBeAg negative and from anti-HBe negative to anti-HBe positive. After seroconversion to anti-HBe positive, most patients have a low level of both viremia and inflammatory liver disease, which is a good prognosis (Gunther et al., 1999). In a minority of patients though, liver disease continues despite seroconversion.

Patients who have endured long-standing liver disease are at high risk of developing end stage liver disease (liver cirrhosis) or heptocellular carcinoma. The pathogenesis of liver cirrhosis is still poorly understood (Grob, 1998). One driving force of heptocellular carcinoma is the random integration of HBV into the host genome. By chance, HBV integration can lead to activation of critical gene promoters of the host (Grob, 1998). Patients with so-called silent HBV infection lack any serological HBV marker (HBsAg, HBeAg, anti-HBe, anti-HBs, and anti-HBc) (Liang et al., 1991).

The variable natural courses of HBV infection (acute self-limited, fulminant, chronic, silent) as well as the different stages (HBsAg positive/negative, HBeAg positive/negative, active or inactive liver disease) and outcomes (remission, heptocellular carcinoma, liver cirrhosis) of chronic infection raise questions as to the clinical relevance of the genetic variability of HBV (Gunther et al., 1999). Are specific variants associated with specific courses or stages of infection?

Symptoms of the onset of HBV infection include abdominal discomfort, nausea, vomiting, and often progresses to jaundice (Harrison, 1999). The prevalence of hepatitis B infection varies geographically. Areas of hyper-endemicity (30% of population infected) include parts of South East Asia and Oceania. In comparison, in North America and much of Europe, only about 0.1% of the population are carriers (see Table 2.3 in Chapter 2).

1.1.3 Ways to combat HBV infection

1.1.3.1 Interferon treatment therapy

Treatment with interferon-α can result in viral elimination in 30-60% of patients (Grob, 1998). Interferon treatment works best in adults who are chronic carriers of HBV who have both low HBV DNA and high alanine aminotransferase (ALT) levels. Interferon-α and interferon-γ successfully eliminates HBV by inducing expression of HLA class I and class II antigens on the surface of HBV-infected hepatocytes (Franco et al., 1988).

A typical course of interferon treatment would be 5MU/m^2 of body surface, administered subcutaneously 3 times weekly for 20 weeks (Sangfelt et al., 1997). Some mild side-effects may occur, such as flu like symptoms for the initial 1-2 weeks of treatment. But more serious side effects such as fatigue, depression and the blood disorder neutropenia may occur, if so, the amount of interferon administered can be reduced by 50% (Sangfelt et al., 1997).

In most cases during interferon treatment, patients will respond with decreased HBV DNA titres, but this does not necessarily last after the treatment has finished. Interferon treatment is less likely to be of benefit if the carrier is of East Asian origin (Sangfelt et al., 1997). Whether this difference in treatment response is due to genetic differences of the host, HBV genotype (East Asian carriers frequently carry HBV genotypes B and C) or to more frequent perinatal HBV infection among East Asian populations has not been elucidated (Sangfelt et al., 1997).

1.1.3.2 Vaccine

The plasma vaccine has been available since the 1980s. But the breakthrough in the prevention of HBV transmission occurred when the genetically engineered hepatitis vaccine, made with viral envelope proteins, was devised (Gunther et al., 1999). The World Health Organisation has been promoting the universal use of the hepatitis B

vaccine since 1991 through the Expanded Programme on Immunisation. In general the HBV vaccine is very successful, and is recognised as the first anti-cancer vaccine. Ideally the vaccine is administered at infancy especially in populations where HBV is endemic (www.who.int/inf-fs/en/fact204.html, 2000).

1.2 Introduction to the Human Leukocyte Antigen complex

The outcome of HBV infection is affected by the host immunological response, such as the Major Histocompatibility Complex (MHC). The human MHC is known as the Human Leukocyte Antigen (HLA) region. The HLA region spans approximately 3.6 megabases on the short arm of chromosome 6 (The_MHC_sequencing_consortium, 1999). The MHC is extremely diverse and with more than 30 alleles at HLA-A and over 50 alleles at HLA-B, it is unequalled by any other complex in the human genome. The function of the MHC molecules is to bind a peptide derived from endogenous or exogenous proteins by specific proteolytic degradation, into a position so they can be recognised by the T-cell receptor (Bodmer, 1995). There are 3 major classes of HLA genes, class I, class II and class III. MHC class I and class II genes have the most numerous and divergent alleles (Parham and Ohta, 1996). Class III consists of factor B and the tumour necrosis factor genes (Parham and Ohta, 1996) and will not be discussed further.

1.2.1 Function of immune cells

The function of the cells of the immune system (B and T cells) is to counter adversity by suppressing infection and maintaining tissue integrity. The specificity, repertoire and memory of the immune response are determined by families of antigen binding molecules: immunoglobins, T cell receptors (TCR's), and MHC class I and class II glycoproteins. Each individual B and T cell has a single antigen specificity, but as a population the cells of the immune system possess great variation (Parham and Ohta, 1996). TCR's recognise short peptide fragments bound by polymorphic MHC glycoproteins. There are two groups of T helper (Th) cells, Th1 and Th2. Th1 cells mainly produce interleukin (IL), tumour necrosis factor (TNF) and interferon-γ (IFN-γ), and favour cellular immune responses

such as macrophage activation (Caillat-Zucman et al., 1998). Th2 cells mainly produce cytokines that stimulate the humoral response (Caillat-Zucman et al., 1998). Th1 and Th2 responses are antagonistic as they suppress one another.

The function of the class I HLA genes is to process and present foreign particles in a complex on the infected cell surface, so that the T cell receptors of the cytotoxic CD8-T-cell recognise them and respond (Bodmer, 1995; Stern et al., 1994). This process ensures all infected cells are identifiable and different from non-infected cells. This recognition system has lead to the selection of a wide variety of mechanisms to escape immune detection by pathogens. HBV is no exception, for example HBV has the potential to down-regulate the expression of surface HLA class I molecules (Bodmer, 1995). Class I molecules bind to a small range of peptides compared to class II molecules, thus are more limited in their response.

Class II molecules are expressed constitutively at high levels on specialised antigen presenting cells, especially dendritic cells, where their function is to present peptides from exogenous proteins to T cell receptors on CD4 positive cells (Bodmer, 1995). The CD4+ cells then modulate the CD8+ cytotoxic T cell lymphocyte (CTL) response and are crucial to the production of neutralising antibodies (Thio et al., 1999).

1.2.2 HLA and HBV

In vitro studies show that HBsAg and HBcAg expressing cells serve as targets for MHC class I–restricted CTLs that are generated from both acute and chronic hepatitis B patients (Tsai et al., 1996). Acute exacerbation of chronic hepatitis B is accompanied by increased T cell responses to HBcAg and HBeAg, but not HBsAg (Tsai et al., 1996).

Viral oligopeptides of 8-15 amino acids are loaded on host cell MHC-class I molecules and are transported to the cell surface. Thus HBV-specific T lymphocytes are able to detect infected cells and destroy them (Grob, 1998). In chronically infected cells, HBV may become partly cytopathogenic – a process still poorly understood – and the viral

DNA may be integrated into the host cell DNA (through a viral transcriptase). If integration leads to activation of certain host genes a heptocellular carcinoma results (Grob, 1998).

A significant proportion of people vaccinated with HBsAg vaccines, particularly hemodialysed patients, either fail to respond or develop very low antibody titres (Caillat-Zucman et al., 1998). The humoral response (stimulated by Th 2 cells) to HBsAg vaccine is influenced by class II allelic variations, which differ in their capacity to bind and present peptides to T lymphocytes (Caillat-Zucman et al., 1998). For example positive responses to the HBsAg vaccine have been associated with DRB1*01 and DRB1*15 alleles, but not associated with the DRB1*03 and DRB1*14 alleles in French Caucasian patients that are hemodialysed (Caillat-Zucman et al., 1998). Moreover some MHC class II molecules may preferentially activate one type of T helper cell over the other. If a Th1 response (cellular) is preferentially induced, the Th2 response (humoral) is inhibited, resulting in the limited production of anti-HBs antibodies (Caillat-Zucman et al., 1998).

1.2.3 Linkage disequilibrium

The HLA system is a powerful anthropological and genetic tool for studying aspects of human biology not only because of the extreme polymorphism at each locus, but also because of the close linkage between loci (Serjeantson et al., 1982). When alleles at closely linked loci occur jointly in individuals more often than would be expected by chance, the population is considered to be in linkage disequilibrium (Serjeantson et al., 1982). Linkage disequilibrium can arise from founder effects, recent mutations, or can be created by population fusion even if there is no disequilibrium in either of the original fusing groups (Serjeantson et al., 1982).

Individuals inherit from each parent an HLA-A and HLA-B gene in the form of a haplotype. For example two copies of the HLA-A gene and two copies of the HLA-B gene are inherited, therefore there are four potential haplotypes. When certain haplotypes

are found more frequently together than would be expected from the known frequencies of each allele it is also known as linkage disequilibrium (Henry et al., 1994).

1.2.4 HLA and Pacific

Polynesian populations are relatively homogenous genetically because they moved into unoccupied areas of the Pacific, and for some 3,200 years they were less affected by admixture with other populations (Murray-McIntosh et al., 1998). Due to their geographic and migration history, Maori and Polynesians show limited diversity in their haplotypes and genes in general (Serjeantson et al., 1982). For example, the New Zealand Maori population has a uniquely low frequency of HLA-A9 (HLA-A*23 and HLA-A*24). New Zealand Maori make up just over 14% of the New Zealand population (StatisticsNZ, 1997) see Table 1.2 for a more detailed look at the ethnicity make up of New Zealand.

Table 1.2 Ethnicity for Population Resident in New Zealand in 1991 and 1996.

Ethnicity	1991 number	1991 percent	1996 number	1996 percent
NZ European	2,488,512	73.75	2,116,332	58.49
other European	169,107	5.01	478,362	13.22
NZ Maori	434,847	12.89	523,371	14.46
Samoan	76,899	2.28	83,718	2.32
other Pacific Island	76,035	2.26	89,466	2.46
South East Asian	16,488	0.49	24,078	0.67
Indian	28,980	0.89	38,403	1.06
Chinese	40,158	1.19	70,227	1.94
other Asian	8,439	0.25	27,975	0.78
Other	6,342	0.19	14,667	0.41
Not Specified/unidentifiable	28,113	0.83	151,716	4.19
Total NZ resident population	3,373,920	100	3,618,303	100

Source: modified from 1996 New Zealand Census of Population and Dwellings.

Presentation of HBV antigens via HLA class I and class II molecules is central to the immune response, therefore differences within HLA alleles may have an effect on HBV

clearance (Thio et al., 1999). HBV disease is also dependent on other host factors such as age at infection. This is a major difference between HBV infection seen in developing countries (mostly perinatal/young) and developed countries (mostly adult). Certain Class II alleles which are associated with HBV persistence have been reported from patients from diverse countries such as West Africa (DRB1*1302) (Thursz et al., 1997), America (DQA1*0501, DQB1*0301) (Thio et al., 1999), and Germany (DQA1*0501) (Hohler et al., 1997). Greater HLA heterozygosity is advantageous as this allows greater numbers of possible HLA-viral antigen combinations (Thio et al., 1999).

Class I alleles have not been significantly associated with HBV persistence, though Thio *et. al.* (1999) reported that 6 class I alleles (HLA-A*2301, HLA-A*3402, HLA-B*4901, HLA-B*4501, HLA-B*4000, and HLA-Cw*1701) were more common in subjects with HBV persistence. No association with homozygosity for class I alleles has been detected.

In a study in which class I and class II MHC antigens were compared in patients with chronic hepatitis B and a healthy control group, the HLA class II alleles DRB1*1301 and 1302 are associated with protection from chronic hepatitis B in Caucasian patients (Hohler et al., 1997). To confirm their findings Hohler *et.al.* (1997) performed a second study in a group of subjects who had recovered from acute hepatitis B infection. The protective effect of the DRB1*1301-02 allele was confirmed in the second study, with 33.3% of the patients who cleared HBV infection being positive for the DRB1*1301-02 allele (Hohler et al., 1997).

HLA-A*0201 is one of the most frequent class I alleles; A*0201 accounts for ~95% of Caucasian A2 positive individuals (del Guercio et al., 1995). A significant percentage of A2 individuals in African, American Indian and Chinese populations express subtypes other than A*0201. In particular A*0202 and A*0205 are frequent in African populations, whereas A*0206 and A*0207 are frequent in Chinese populations (del Guercio et al., 1995). HLA-A*2402 is the most common HLA class I allele in East Asia (Tanaka and Takiguch, 2001) a region in which there are approximately 200 million

hepatitis B virus carriers (Tanaka and Takiguch, 2001). The HLA-A*02 allele family is the most frequent allele family in the New Zealand Maori population (Henry et al., 1994).

Henry *et.al.* (1994) determined the class I gene, antigen and haplotype frequencies for New Zealand Maori and New Zealand Europeans. Their results are summarised later in Table 4.24 of the results chapter with a summary of HLA results for the Kawerau samples that are the subject of this thesis.

1.3 Aims of Thesis

To understand the complicated interactions between the HBV and it's human host, investigations that follow the progress of the viral infection are vital. Longitudinal studies look at what changes, if any, have occurred in the virus over a specified time period. By carrying out such studies, it is hoped that a greater understanding of the virus will be gained. Changes in the disease, such as acute to chronic hepatitis may be narrowed to specific mutations which the virus gains. The aim of this study was to compare viral genome sequences from 1985 to viral sequences from the same host from 1998.

To fully understand the viral:host interaction, genetic data from the host also needs to be examined. HLA typing was performed on the individual hosts. The aim was to compare the hosts that had different clinical types of HBV infection to see if the presence of any particular HLA allele was correlated with a particular viral clinical type. The HLA data from individuals with HBV infection were compared to the general New Zealand population.

Chapter 2 Characteristics of Viral Genome

The genetic factors of the hepatitis B virus influence the expression of viral antigens (Lindh et al., 1999), and these are what the host's immune system responds to. Liver damage in HBV infection is thought to result from the response of the host's immune system to the expression of viral antigens, and not from a cytopathic effect of the virus itself. Many factors influence the host's immunological response such as age and health of host at infection, genetic host factors (such as HLA type), and genetic variability of the virus. The genetic variability of the virus, and the genetic factors which influence the expression of viral antigens and the virus itself will be discussed in this chapter.

The hepatitis B virus (HBV) genome consists of four overlapping open reading frames encoding the core protein, the X gene region, reverse transcriptase (or polymerase), and the surface protein (refer to Figure 1.1). Expression of these genes is regulated by four viral promoters which produce four HBV-specific RNAs (Chen et al., 1993). The biology of each of these will be discussed.

2.1 Pre core / Core Gene

The hepatitis B e-antigen (HBeAg) (encoded by the pre-core (preC) and core (C) gene region of hepatitis B Virus) has clinical and epidemiological significance, because patients that express HBeAg tend to have more severe liver disease and are more infectious than those that do not express HBeAg (Horikita et al., 1994). In general, the presence of HBeAg in the serum correlates with high viral titre, high levels of viral replication and elevated alanine aminotransferase activity. Correspondingly, HBeAg negative patients usually have low viral titres, low levels of hepatitis B virus replication and normal alanine transferase activity (Horikita et al., 1994; Lai et al., 1994).

Seroconversion from HBeAg positive to HBeAg negative is usually due to mutations in the preC region that prevent the synthesis of HBeAg. The most frequent mutation is a point mutation from G→A at nucleotide 1896 in the preC region converting codon 28 from

tryptophan (TGG) to a stop codon (TAG) (Arauz-Ruiz et al., 1997; Horikita et al., 1994). It is especially common in the Mediterranean area (genotype D) and in the Far East (genotypes B and C) and is found in patients with severe, rapidly-progressing chronic active hepatitis, positive for HBsAg and HBV DNA, although negative for HBeAg (Lindh et al., 1999; Zhang et al., 1996). The 1896 mutation functions as an immune escape mutant, since it prevents the expression of HBeAg on the surface of HBV infected hepatocytes, and therefore protects these from being eliminated by cytotoxic T cells specific for HBeAg (Arauz-Ruiz et al., 1997). Mutants of an HBeAg negative phenotype progressively replace the non-defective precore HBV populations (HBeAg positive phenotype) as the mutants are selected for by immune pressure in persistently infected hosts (Horikita et al., 1994; Okamoto et al., 1990).

Mutations of the virus can affect the clinical status of the host. Obviously mutants with a less negative influence on the longevity of their hosts would manage to survive, but only if they manage to transmit to a new host. Mutants that have a significant pathogenic affects on hosts, however, can arise and induce severe disease in their host. HBV mutants of a HBeAg negative phenotype are typical examples. There are two clinical profiles of carriers who seroconvert to anti-HBe; some clear the hepatitis infection while others develop severe hepatitis leading to fulminant disease (Horikita et al., 1994).

Both clinical profiles have precore mutations terminating HBeAg expression. Therefore other factors are thought to determine clinical outcomes. These include factors that may affect replication activity of the mutants, which in turn might influence disease inducing activity, or induce amino acid substitutions for altered recognition by ccll-mediated immunity of hosts. This view is supported by anti-HBe positive active hepatitis with high serum levels of HBV DNA (Horikita et al., 1994) and unique mutations clustering in the putative T cell epitopes in the HBV core protein in chronic active hepatitis (Horikita et al., 1994).

Having the TGG $\rightarrow$ TAG mutation at codon 28 in the preC region is the most common mutation for a HBeAg negative phenotype. Other mutations in HBeAg negative

phenotypes are likely to be within the X gene, where the pre C/C promoter also lies (nucleotides 1742-1849).

The limited data available from previous longitudinal studies suggest that most of the mutations appear during periods of increased immune pressure (Bozkaya et al., 1996). New changes in the HBV core gene nucleotides and amino acid sequences are rarely detected in patients who are still in the immune tolerance phase (patients who remain HBeAg positive with normal amino-transferase levels) (Bozkaya et al., 1997). However very high rates of change can be found during the immune clearance phase (HBeAg positive patients with elevated amino-transferase levels, or HBeAg negative patients) (Bozkaya et al., 1996). Interferon therapy does not induce a high rate or a specific pattern of HBV core gene mutations (Bozkaya et al., 1996).

Amino acid changes in the HLA-A2 restricted CTL epitope (amino acid residues 18-27 of core gene, nucleotides 1952-1981), predominantly isoleucine to valine substitutions at position 27, occur with similar frequencies irrespective of HBeAg seroconversion, interferon treatment, and response to interferon therapy (Bozkaya et al., 1996). Mutations at amino acid positions 21 and/or 27 may affect T cell recognition of the HLA-A2 restricted cytotoxic T cell (CTL) epitope (Bozkaya et al., 1996).

When compared to patients with similar clinical profile it was found that interferon therapy (see section 1.1.3.1 of Chapter 1) did not induce a higher rate of HBV core gene mutations, and that non-responsiveness to interferon therapy could not be accounted for by an increased number or specific patterns of HBV core gene mutations (Bozkaya et al., 1996).

Either progression of liver disease or clinical remission can occur after seroconversion to anti-HBe positive. Anti-HBe positive progressive hepatitis is particularly common among Mediterranean peoples, indicating that there may be a host genetic factor involved (Carmen et al., 1997).

There are associations between virus clearance during acute infection and the presence of specific HLA class II alleles. Mutations in hepatitis B core protein have been related to disease progression (Carmen et al., 1997). Soon after the appearance of pre-core mutants there is an initial burst of core protein amino acid substitution. Loss of HBeAg leads to a reduced immune response, and thus selection pressure against the core protein (Carmen et al., 1997). Mutant strains are escaping from the immune response against the core protein rather than causing peaks of hepatitis (Carmen et al., 1997).

Patients who went into remission have mutations in the Th cell epitope of the virus, whereas patients with ongoing disease have mutations which occur in B cell epitopes (Carmen et al., 1997).

The viral encapsidation signal spans both the precore and core genes (nt 1852-1930) of the pre-genome. The primary functions of the mutations in the precore region below is to enhance stability of the secondary structure to ensure perpetuation of viral replication (Lok et al., 1994). The four major mis-sense mutations (M1 to M4) are found in the precore region that have an effect on the secondary structure of the DNA (Lok et al., 1994) are:

- M1, C to T at 1856; proline to serine at codon 15.
- M2, G to A at 1896; tryptophan to STOP at codon 28.
- M3, G to A at 1898; glycine to serine at codon 29.
- M4, G to A at 1899; glycine to aspartate at codon 29.

The commonest conserved mutation is M5 T to C at 1858; proline to proline at codon 15. See Figure 2.1. The primary function of the mutations in the pre-core region is to enhance stability of this secondary structure to ensure the perpetuation of viral replication. The pre-genome encapsidation signal serves as a template for the initiation of reverse transcription from pre-genomic RNA to the minus strand DNA (Lok et al., 1994).

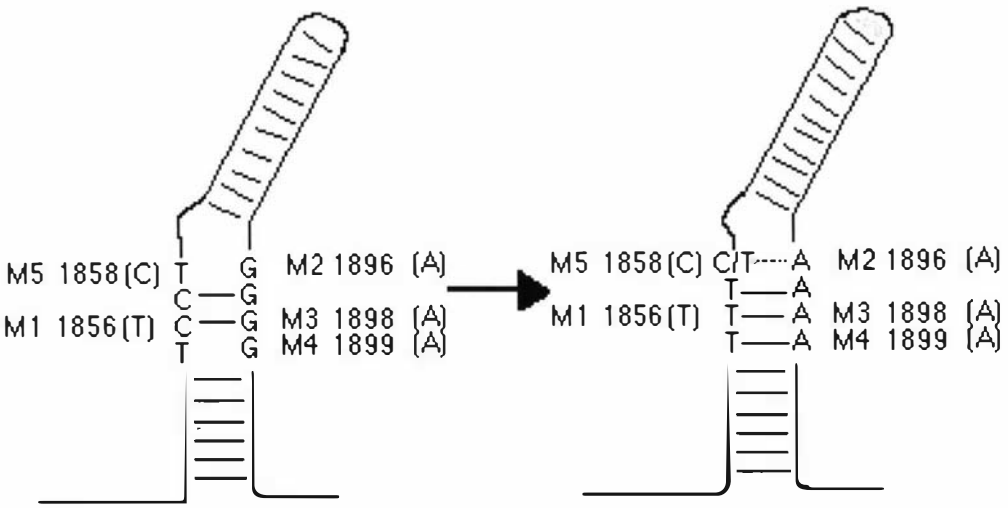

Figure 2.1 Schematic diagram showing positions of mis-sense mutations (M1 to M4) and conserved mutation (M5).

M1 and M2 are mutually exclusive, M3 is only associated with M1 (50% of patients with M1 had M3), M4 predominantly is associated with M2. All patients with M1 have M5, but none with M2 had M5. M2 and M4 enhance stability of a secondary structure stem by providing two additional paired sites, M1 destroys an existing base pair, however M1 only occurs in presence of M5 which provides an extra paired site. Lok *et.al.*'s (1994) data supports the proposed secondary structure of the pre-genome encapsidation sequence.

HBV specific cytotoxic T lymphocytes (CTL) play a major role in the control of viral infections through their ability to identify and neutralise virus-infected cells. This is accomplished by recognition of short viral peptides being expressed at the cell surface in the binding groove of HLA class I molecules (Bertoletti et al., 1997). CTLs also release cytokines at the site of infection. The relative strength of the CTL response is a crucial determinant of the final outcome of HBV infection, and a number of CTL epitopes have already been identified within the different HBV antigens. Among them, the amino acid sequence 18-27 of HBcAg has been shown to be widely recognised by CTLs of HLA-A2-

postive patients with acute HBV infection, and is known as the CTL epitope. The main contribution to HLA binding of HBc 18-27 sequence was given by residues 19 and 27 (positions 2 and 10 of the peptide) with a significant role also played by residue 24 (position 7) (Bertoletti et al., 1997). The CTL epitope is also able to activate a specific CTL response during chronic hepatitis B, although the CTL activity at this stage of infection is generally weaker (Bertoletti et al., 1997). Patients who successfully recover from acute viral hepatitis develop strong and multi-specific HLA class I and class II restricted T-cell responses, whereas these responses are weak or absent in patients with chronic hepatitis B (Bertoletti et al., 1997).

Amino acid substitutions that define the different HLA-A2 subtypes can potentially influence not only the binding, but also the conformation of the peptide/HLA complex, and recognition by the T cell receptor (Bertoletti et al., 1997). The core protein CTL epitope can bind different HLA-A2 subtypes, namely A*0201, A*0202, A*0205 and A*0206 (Bertoletti et al., 1997). HLA-A*0202 and HLA-A*0205 are frequent among African individuals; A*0206 is frequent among the Chinese population. The hepatitis B virus carrier rate for these populations is from 5 to 30 %. In patients with a particular genetic background, this region thus has the potential to activate both HLA class I and HLA class II restricted HBcAg negative specific T-cell responses (Bertoletti et al., 1997).

The double mutation (AGG to TGA at nucleotides 1762-1764, changing amino acids lysine (130) to methionine and valine (131) to isoleucine) in the part of the X open reading frame that contains the pre-core/core promoter has been linked to HBeAg negative phenotype and has been associated with severe liver damage (Hannoun et al., 2000; Lindh et al., 1999). This double mutation has also been observed in the HBeAg positive stage (Lindh et al., 1999; Takahashi et al., 1995). In these situations the HBeAg is e-suppressed rather than e-negative.

2.2 The X Region

Although its exact role in the viral life cycle is not clear, it has been shown that the X protein is essential for viral replication *in vivo*. Several functions have been attributed to the HBx protein, see Table 2.1 for details. The X gene region spans from nucleotides 1376 to 1840.

The HBV X protein is thought to be involved in the development of HBV-associated hepatocellular carcinoma. The X protein has been shown to transactivate several viral and cellular targets, including oncogenes such as c-fos and c-myc (Becker et al., 1998). Transcription of the X-gene appears to be necessary for the trans-activating function of the X protein (Twu and Schloemer, 1987). Transactivation by the X protein depends on at least two distinct cellular DNA-binding transcription factors, AP-1 and AP-2 (Seto, 1990). The functional significance of these protein-protein interactions are not fully understood. The X protein has been found in both the nucleus and cytoplasm of transfected cells, indicating that the protein may interact with cellular proteins localised in both these compartments (Becker et al., 1998).

Expression of the X protein interferes with the ability of the cell to repair damaged DNA (Becker et al., 1998). Becker hypothesised that the inhibition of DNA repair by HBx may be mediated by its ability to bind with X-associated protein 1 (XAP-1), a DNA repair protein. XAP-1 shares 99% homology with the simian UV-damaged DNA binding protein (UVDDB). The results of Beckers study demonstrate that HBx expression is associated with an approximately 45% decrease in DNA repair capacity, as measured by two independent DNA repair assays (Becker et al., 1998).

Table 2.1: Reported functions and protein associations of the X protein.

Function / Protein Association	**Reference**
Ribo-deoxy ATPase activity	(De-Medina et al., 1994)
Activates Cellular protein kinase C signalling	(Kekule et al., 1993)
Activates *Ras-Raf-Map* kinase pathways	(Benn and Schneider, 1994; Cross et al., 1993; Natoli et al., 1994)
Binds CREB and ATF2	(Maguire et al., 1991)
Binds the protease tryptase TL2	(Takada et al., 1994)
Binds p53 (tumour suppresser protein)	(Truant et al., 1995)
Binds TATA-binding protein	(Qadri et al., 1995)
Binds a regulatory α subunit of a proteosome complex	(Huang et al., 1996)
Binds RNA polymerase subunit RBP5	(Becker et al., 1998)
Interacts with X AP-1 protein	(Lee et al., 1995; Sitterlin et al., 1997)
Acts through both AP-1 and AP-2 sites	(Seto, 1990)

The region of the X protein required to interact with XAP-1/UVDDB are amino acids 55-101 (nucleotides 1538-1677). Sitterlin *et.al.* (1997) found a correlation between the ability of the X protein to bind UVDDB and the ability of the X protein to transactivate an AP1-containing promoter–enhancer-driven reporter construct. Mapping results obtained with X protein point mutations are also consistent with the hypothesis that UVDDB binding may be important for X protein transactivation (Becker et al., 1998).

The transactivating function resides in a 944 nucleotide *Eco*RV-*Bgl*II DNA fragment (1040-1984) of the HBV genome that contains the X structural gene and its promoter element. Removal of the promoter from the X structural gene results in loss of the transactivating function. A frame-shift mutation within the X gene region also eliminates the transactivating activity, suggesting that the X antigen could play a role in HBV infection by activating the expression of the host genes (Twu and Schloemer, 1987).

A transactivating product of the X gene region is responsible for the activation of the regulatory region that governs the expression of a eukaryote gene, the human beta-interferon gene (Twu and Schloemer, 1987).

The X gene region not only encodes the X protein, but also contains many of the functionally important elements such as the core promoter, enhancer II and two direct repeats, which are implicated in the replication and expression of HBV DNA (Uchida et al., 1997). The core promoter and enhancer II sequences widely overlap each other, the former is located at nt 1591 to 1851, the latter at nt 1687 to 1805 (Uchida et al., 1997). Enhancer I is located outside the X region at nt 966 to 1308. Both enhancer I and enhancer II affect the activity of all the core/pregenomic, pre SI and pre SII/S and X promoters with varying activity (Uchida et al., 1997). A negative response element (NRE) upstream of enhancer II (nt 1613 to 1636) represses the activity of enhancer II. Mutational analysis revealed that nt 1616 to 1621 are essential for the repression of the enhancer (Lo and Ting, 1994). The discovery of the NRE indicates that HBV gene transcripts are controlled by both positive and negative regulation, adding to HBV complexity.

2.3 Polymerase Transcript (P gene)

The P gene transcript is 3.5kb long and starts from the core promoter. The precore/core genes and P gene are expressed independently and act in trans, therefore the HBV polymerase gene product is synthesised independently from the HBV core gene (Radziwill, 1990). The P protein consists of 3 functional domains and a non-essential spacer arranged in the following order: terminal protein, spacer, reverse transcriptase/DNA polymerase, and Ribonuclease H. The spacer region can be deleted to a large extent without significant loss of endogenous polymerase activity (Radziwill, 1990). Except for the variable spacer region, the hepadnavirus P protein sequence is conserved, even between the distantly related avian and mammalian members of this virus family (for example HBV and DHBV) (Radziwill, 1990).

Polymerase and RNase H protein sequences are highly conserved, except for the second half of the DNA polymerase domain, which is poorly conserved between mammalian hepadnaviruses (Radziwill, 1990)). The conserved zones are characterised by segments of significant sequence homology to retrovirus DNA polymerase/reverse transcriptase and

also to RNase H (Radziwill, 1990). The terminal protein domain is unique to the hepadnavirus family. The function of these domains (terminal protein, polymerase, RNase H) are to catalyse the major steps in hepadnavirus reverse transcription i.e priming, DNA synthesis, and removal off the RNA template (Radziwill, 1990).

The reverse transcriptase domain contains the YMDD amino acid sequence motif that is present in many retrovirus reverse transcriptase regions. It is essential for retrovirus activity in the HBV, it is situated at amino acids 538-541 of the polymerase region (Radziwill, 1990).

Hepadnavirsuses are characterised by a small circular DNA genome of about 3.2 kilobases which is gapped in one DNA strand (the plus strand) and contains a terminal protein linked to the 5' end of the complementary DNA minus strand. This peculiar structure arises intracellularly by reverse transcription of an RNA pregenome through a process which apparently involves the terminal protein during priming of DNA synthesis and which then, by analogy to retroviruses and retroelements, is assumed to be catalysed by the concerted action of DNA polymerase/reverse transcriptase and RNase H activities (Radziwill, 1990).

In comparison to retroviral polymerase genes, hepadnaviruses lack an endonuclease, which is not surprising as there is no specific integration into the host genome. Also the retroviral protease is replaced by the HBV DNA terminal protein (Radziwill, 1990).

Encapsidation of pregenomic RNA into nucleocapsids is a selective process which depends on specific RNA-protein interactions (Bartenschlarger, 1990). The P gene product containing all functional domains is required for the encapsidation of HBV pregenomic RNA (Bartenschlarger, 1990). The P protein is required as a structural component for packaging of nonviral RNAs fused to the HBV encapsidation signal. The P protein acts primarily in *cis* as the pre-genomic RNAs from which P protein is synthesised are preferentially encapsidated (Bartenschlarger, 1990).

Despite containing a DNA genome, hepadnaviruses resemble retroviruses in that they replicate via reverse transcription of an RNA pregenome transcript. This reaction takes place in a particle which is composed of core proteins, P proteins and the RNA pregenome. Thus, packaging of this RNA is an essential step leading to the formation of replication-competent nucleocapsids (Bartenschlarger, 1990).

In the polymerase region, where the intergenotypic variability is low and where active sites for reverse transcriptase/polymerase and RNase H functions reside, the mutation frequency is lower – reflecting the fact that mutations in this region are likely to be deleterious to the virus (Hannoun et al., 2000).

2.4 Surface Genes

The HBV virion envelope contains three related surface glyco-proteins. These proteins are encoded inframe from a single open reading frame and share a common stop codon. They are preS1 – large protein; preS2 – middle protein; and S – small protein (Ryu et al., 2000). It is within these three genes that all the envelope epitopes which determine the subtypes of the surface protein reside.

These envelope proteins have epitopes that the immune defence mechanism responds to, eliciting virus-neutralising and protective antibodies. In the case of preS1, antibodies to the peptide in PreS1 (aa 21-47, nt 2910-2990) were shown to neutralise HBV infection in chimpanzees. Amino acids 21-47 of preS1 are the putative receptor binding region (Hannoun et al., 2000). It contains a specific binding site for a human hepatocyte receptor (Ryu et al., 2000). Mutations within this region result in mild liver damage, as binding affinity is reduced (Hannoun et al., 2000). The reduced binding affinity causes a diminished immune response, which results in less severe liver damage as it is the host immune response to the virus which causes the damage to the liver cells. Thus monoclonal antibodies that bind to amino acids 21-47 of HBV preS1 would be useful not only in elucidating the function of preS1 in viral infectivity and the process of neutralisation, but

also in the development of therapeutic reagents for the preventative treatment against HBV infection (Ryu et al., 2000).

Cells infected with the HBV virus produce both infectious and non-infectious virus particles. The S protein is found in large quantities on both particle types, whereas the L protein is preferentially localised on infectious viral particles (Ryu et al., 2000). The L protein is speculated to have a major role in viral infectivity and assembly (Ryu et al., 2000).

Unfortunately, chronic HBV infection is associated with T cell hypo-responsiveness or tolerance. Several approaches to overcome CTL tolerance using transgenic mice-model systems, have been investigated. Targeting subdominant epitopes (such as Env.364 and Env. 362 instead of Env. 28) may be an effective way to overcome CTL tolerance (Sette et al., 2001).

In the preSI region at nucleotides 2854-2886 (aa 1-11) there is a variable region which is genotype specific, for example it is deleted for genotype D but not for genotype C (Hannoun et al., 2000). The region does not contain any areas of significance, so can endure the deletion.

2.4.1 Serotypes and Genotypes

The S region has mutations at sites linked to *w/r* seroreactivity, these are T(126) → A; I (110) →L; Y(161) →F. Also N-→S at amino acid 40 has recently been reported as being frequent in Japanese *adw* carriers with hepatocellular carcinoma (Hannoun et al., 2000).

Subtypes of the hepatitis B surface antigen (HBsAg) have been defined by two mutually exclusive determinant pairs, *d/y* (Le Bouvier, 1971) and *w/r*(Bancroft et al., 1972), and a common determinant *a.* These subtypes are *adw, ayw, adr, ayr.* These four major subtypes were further subdivided resulting in 9 different subtypes being identified. The

subtypes are *ayw1, ayw2, ayw3, ayw4, ayr, adw2, adw4, adrq-* and *adrq+* (Courouce-Pauty et al., 1978).

The *ayw* subtype is found among drug addicts world-wide and is dominant among the carriers in north-western Europe. Subtypes with the *r* determinant were exclusively confined to populations of the Far East.

A study performed on two blood donors carrying surface antigens of compound subtypes *adyr* and *adwr* respectively, showed that the amino acid substitutions at positions 122 and 160 alone explained the expression of *d/y* and *w/r* specificity, respectively (Magnius and Norder, 1995; Okamoto et al., 1987b). Both the *d* to *y* and *w* to *r* changes were mediated by a shift from lysine to arginine at the corresponding positions (Magnius and Norder, 1995; Okamoto et al., 1987b).

Sequencing of the viral genomes in the late '80s produced a genetic classification based on the comparison of complete genomes. At present there are seven genotypes identified, A to G (Magnius and Norder, 1995; Stuyver et al., 2000), they are defined by a divergence of greater than 8% between each genotype, with genotype F being the most divergent at 14% (Magnius and Norder, 1995). Table 2.2 gives a summary of which serological subtypes belong to which genotype and their geographical distribution. Perinatal transfer and adult drug mis-use are the primary cause of transfer of the infection. The European strain of genotype D (*ayw3*) is known to be associated with intravenous drug abuse.

Table 2.2 Geographical distribution of HBV genotypes and subtypes.

Genotype	Subtype	Areas of high prevalence	Carrier %
A	*adw2*	North-western Europe	0.01
	ayw1	Central Africa	5-20
B	*adw2*	Indonesia, China	5-30
	ayw1	Vietnam	5-30
C	*adw2*	East Asia	5-30
	adrq+	Korea, China, Japan	5-30
	adrq-	Polynesia	5-20
	ayr	Vietnam	5-30
D	*ayw2*	Mediterranean area	5-20
	ayw3	India, Europe	0.01
E	*ayw4*	West Africa	5-20
F	*adw4q-*	American natives, Polynesia	5-20
G	*adw2*	France, America	0.01

There are characteristic differences observed between each subtype. The only differences between *ayw2* and *ayw3* are found at amino acids 125 and 127, which are Thr and Pro for *ayw2* and Met and Thr for *ayw3*, respectively, though later studies showed some *ayw3* strains do not express Met 125 (Magnius and Norder, 1995). Variations at amino acid 127 explain the *w2-w4* variations and are Pro, Thr and Leu/Ile, respectively. In general, HBV strains within genotype C have an Ile at 126 (Magnius and Norder, 1995). Table 2.3 summaries the amino acid specifying determinants of HBsAg.

Table 2.3 Amino acid residues specifying determinants of HBsAg.

Position	Amino acid	Specificity
122	Lys	*d*
	Arg	*y*
	Arg	*w1*
127	Pro	*w1*
	Pro	*w2*
	Thr	*w3*
	Leu/Ile	*w4*
134	Phe	*w1*
159	Ala	*w1*
160	Lys	*w*
	Arg	*r*

The dominant epitope cluster in HBsAg is within the major hydrophilic region and is termed the *a* determinant. Based on peptide studies it is widely regarded as being between

amino acids 124-147 (nucleotides 526-597). Antibodies to the surface antigen isolated from the sera of vaccinated individuals bind to amino acids 124-147 (Magnius and Norder, 1995).

The long time-course of chronic HBV infection could favour a selection of escape mutants (Zhang et al., 1996). The glycine to arginine mutation at amino acid 145 in the second polypeptide loop of the *a* determinant (AAs 139 to 147) is a vaccine escape mutant, and is another example of the genetic diversity of HBV created by point mutations and selective pressure (Zhang et al., 1996). The DNA sequence encoding the first loop of the *a* determinant is more prone to variation than that of the second loop (Zhang et al., 1996). Prevalence of divergent sequences was highest in HBsAg negative / anti-HBs positive patients with HBV DNA (Zhang et al., 1996). The majority of the patients with HBV epitope mutants must have been recruited from chronic HBsAg carriers with HBeAg or with anti-HBe and precore mutation leading to a translational stop codon (Zhang et al., 1996). Likewise, escape mutants with mutations prohibiting the expression of immunodominant HBsAg epitopes occur in recipients of hepatitis B vaccines (Horikita et al., 1994).

2.4.2 S protein

S antigen is made up of approximately 100 S chains creating a 22nm HBsAg sphere. The S protein is synthesised at membrane-bound ribosomes at the rough endoplasmic reticulum (ER). Its topology is determined during synthesis by signal sequences in the 226 amino acid long peptide chain (Figure 2.2). An N-terminal signal I (Sig I) leads to co-translational translocation of the N-terminus across the ER membrane during translation. This signal contains a hydrophobic stretch between amino acids 8 and 22. It is not cleaved, nor does it contain any anchor function to hold the peptide chain in the membrane. But most models predict that this sequence is embedded in the membrane due to the anchor function of the second downstream signal II (sig II). This signal is within the central part of the S protein, and contains a hydrophobic stretch between amino acids 80-98. It causes the anchoring of the hydrophobic sequence in the membrane and in addition the

translocation of the downstream peptide sequence into the ER lumen. The last 56 amino acids at the C-terminal end of the S protein are very hydrophobic; computer models predict that this region forms two α-helices which cross the ER membrane twice. Experimental evidence suggests that the C-terminal end of the S protein juts out into the ER lumen. The region between sig II and the hydrophobic C-terminus forms a luminal loop which carries the major antigenic determinants of HBsAg (Bruss et al., 1996). This region is also very hydrophilic.

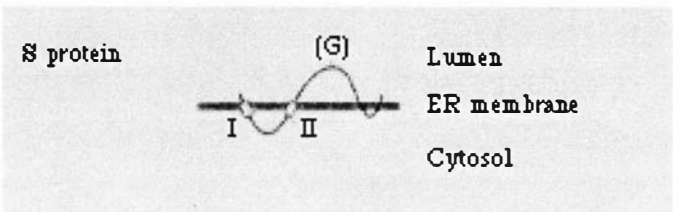

Figure 2.2 The Small protein, showing partial N-glycosylation site (G). Source (Bruss et al., 1996).

The S protein is co-translationally modified, an asparagine at residue 146 in the luminal loop is N-glycosylated (gp=27kD), although this only occurs in about 50% of S proteins (Bruss et al., 1996).

The S protein contains 14 cysteine residues (4 in the cytosolic loop, 1 in the sig II region, 8 in the luminal loop, and 1 close to the C-terminus) which are involved in intra- and intermolecular S-S bridging (Bruss et al., 1996). Shortly after synthesis, the S protein forms dimers which are stabilised through disulphide linkage. The cys residues used for this linkage are in the luminal loop (Bruss et al., 1996). Two dimensional gel electrophoresis demonstrated that glycosylated -S and un-glycosylated -S chains paired with each other without any preference (Bruss et al., 1996).

2.4.3 M protein

M protein differs from S protein by an additional 55 amino acid N-terminal preS2 domain (Figure 2.3). Biosynthesis of the M protein is very similar to the synthesis of the S protein. The preS2 domain is co-translationally translocated into the ER lumen by sig I in its downstream S domain. Asn at residue 4 is N-glycosylated, therefore the 30 kD peptide appears on polyacrylamide gels as a 33 kD protein. The partial S domain glycosylation results in a 36 kD protein. The 30 kD protein is sometimes observed, but is very inefficiently incorporated into subviral particles (Bruss et al., 1996). The M protein forms cys-linked dimers; in fact M and S proteins form mixed dimers without any preference. The preS2 domain contain no cys residues, thus it is the same cys residues from the S domain which form these linkages.

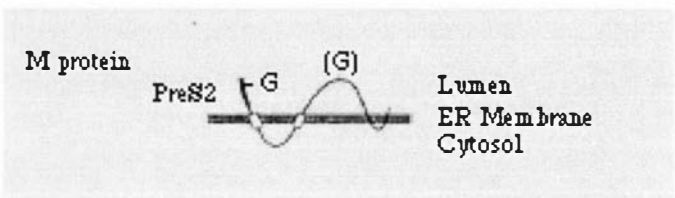

Figure 2.3: The M protein, showing the potential N-glycosylation sites (G). Source: (Bruss et al., 1996).

2.4.4 L protein

The large surface protein (L) adopts two different transmembrane topologies due to a post-translational switch of the folding in approximately half of the L proteins (Figure 2.4). The L molecules which expose their N-terminal preS1 domain on the viral particle surface are most likely ligands for a putative virus receptor, and determine the species specificity and liver tropism of this virus. L molecules with internal preS1 domains are required in virion morphogenesis and mediate contact to the nucleocapsid like a matrix protein (Bruss et al., 1996). Overexpression of this form of the L protein is also responsible for the inhibition of viral particle release (Bruss et al., 1996).

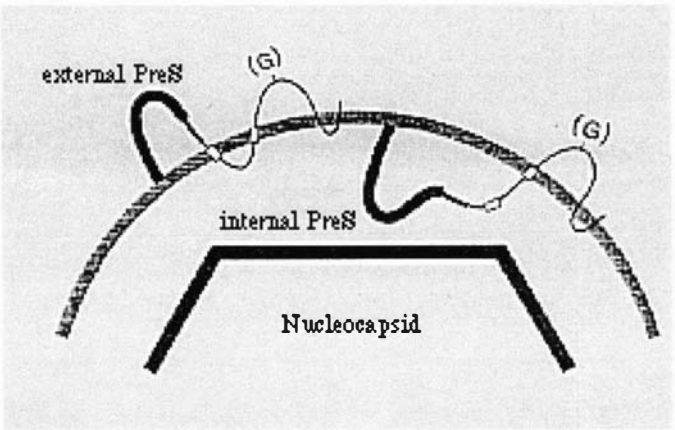

Figure 2.4 Different topologies of the L protein. Source (Bruss et al., 1996).

In the L protein with an internal preS1 domain, the N- glycosylation site in preS2 (Asn 4, which is used in the M protein) and another potential site in preS1 (Asn 15) are not used, in agreement with their cytosolic location (Bruss et al., 1996). The L protein forms mixed cys-linked dimers with M and S proteins without any preference (Bruss et al., 1996). This indicates that the cys residues used for the linkages are within the S domain, therefore this region of all 3 surface proteins is probably folded in the same way (Bruss et al., 1996).

L proteins with external preS domains dispose their preS domains at the outside and have a topology similar to the M protein. The external preS domains are generated by a post-translational mechanism is mainly based on the missing N- glycosylation of Asn 15 in preS1 and Asn 4 in preS 2 of the L protein. When the preS domains are forced to co-translationally enter the ER lumen, these sites are N- glycosylated (Bruss et al., 1996).

One not well understood function of the L protein is that it inhibits the release of subviral particles and virions. This retention function is mediated by N-terminal structures of the L protein (Bruss et al., 1996). Another function of the L protein, is that it binds to the putative virus receptor. Binding studies to hepatocyte membrane preparations suggest that the external preS sequence of the L protein can mediate attachment to the hepatocyte (Pontisso et al., 1989). This is the role of the L proteinin virion formation. The L protein sequence which interacts with the nucleocapsid is probably located between amino acids 103 and 253 (start of sig II) (Bruss et al., 1996). The L protein also plays a part in the

regulation of the copy number of episomal HBV genomes in the nucleus (Bruss et al., 1996).

The dual function of the L protein, being a matrix protein in virion morphogenesis and a ligand for a virus receptor, is reflected by the dual topology of this protein and is divided between the C-terminal and N-terminal part of the preS, respectively (Bruss et al., 1996).

The relative ratio of large surface protein to the middle and small surface proteins is crucial to HBV replication. A feedback mechanism ensures a balanced synthesis between these proteins. This feedback mechanism exists in that the large surface protein activates the S promoter; this activation is correlated with the intracellular retention of L. Therefore, over expression of L and subsequent intracellular particle retention appear to activate the S promoter by an intracellular signalling pathway induced by ER stress. This activation would in turn lead to increased synthesis of M and S proteins restoring the proper ratio of L:M:S proteins, and, hence allow secretion of both subviral and virion particles (Xu et al., 1997).

2.5 Quasi-species Distribution

HBV replicates by means of an RNA intermediate using its own reverse transcriptase. Because this enzyme lacks proof reading activity, a high number of mutant genomes are introduced into the viral population of a host. It has been estimated that at a single nucleotide position of HBV 1.3×10^{-5} mutations occur in an infected individual each year (Carmen and Thomas, 1992; Okamoto et al., 1987a). Thus the virus exists in a quasispecies distribution where a population of viruses that share a common origin but which have variant genome sequences as a result of mutation, drift, and the impact of selection (Smith et al., 1997). Quasispecies distributions have been shown in patients with chronic hepatitis (Brown et al., 1992). However, it must be remembered that the overlapping reading frames in the viral genome limit the biologically fit mutations. Combined, these give HBV a distinctive distribution of mutations.

Chapter 3 Materials and Methods

In 1985 whole blood was taken from 48 patients who were asymptomatic carriers detected in a HBV survey of Kawarau in 1985. The survey was conducted by the Child Health and Hepatitis Foundation, Whakatane. Each patient has since been monitored by the Foundation. In 1998 more whole blood was taken from the same patients. The aim of this project is to investigate, for three clinical types of hepatitis B infection, differences over a 15 year period, and then to determine if a correlation exists between specific HBV variants and particular clinical states.

With all samples, DNA was extracted and then amplified using the polymerase chain reaction (PCR). The resulting PCR fragments were sequenced directly.

3.1 DNA Extractions

Extractions were carried out according to the High Pure™ Viral Nucleic Acid Kit isolation protocol with the following modifications.
In a nuclease-free 1.5 ml centrifuge tube 200 µl of whole blood, 200 µl working solution (carrier RNA-supplemented Binding Buffer), and 1.8M Proteinase K in a final volume of 440µl was incubated at 72°C for 10 minutes, with frequent vortexing.

Post incubation, 100 µl of isopropanol was added and mixed. The entire 540µl was then added to the High Pure™ Viral Nucleic Acid Isolation column and spun at 14,000 rpm for 2 minutes. The column was then washed with 450 µl of wash buffer and centrifuged at 14,000 rpm for 1 minute. This step was repeated and then the column was 'dried' via a quick 10 second spin at 14,000 rpm to remove any residual solution.

The viral DNA was eluted from the column with 50µl of prewarmed milliQ H_2O. To ensure maximum viral DNA elution the column was incubated at 50°C for 10 minutes before elution at 13,000 rpm for 2 minutes. The elution was repeated to ensure maximum recovery of the viral DNA. Viral DNA was stored at -20°C.

3.2 PCR Amplification

The Polymerase Chain Reaction (PCR) is an efficient means of amplifying DNA sequences. First the double stranded DNA template is separated using a high temperature of 94-96°C. This is followed by a rapid drop to 45-60°C for 30-60 seconds to allow the oliogonucleotide primers to anneal to the single stranded template. The annealing temperatures are based on the melting temperatures of the individual primer pair and homogeneity to the template. Higher annealing temperatures allow more specific amplifications. After primer annealing the temperature is increased to 74°C where the polymerase enzyme replicates the DNA template. Each repeat of these cycles amplifies the DNA in an multiplicative manner. DNA contamination is a potential problem with PCR, as only a few picograms (pg) of DNA is needed for each reaction. Consequently controls were used to create DNA sterile conditions and to monitor for contamination. Aerosol resistant tips were used and blank control reactions containing no intentionally added DNA were run to check that all reagents were clean (Palumbi et al., 1991; Saiki et al., 1988). All reactions were done in 0.2ml Scientific Specialties Incorporated research tubes; the PCR machines used were MJ research PTC-200 Peltier Thermal Cycler, Hybaid PCR Sprint Thermal Cycler, and Hybaid PCR Express Thermal Cycler. In all reactions, a heated lid was used so no oil was required. In the MJ research machine block settings were used and in the Hybaid machines probe settings for greater accuracy were used.

For PCR amplifications where the final product was expected to be greater than 2 kb the long range Expand High Fidelity PCR system (Roche) was used as it has greater fidelity than Taq long range enzyme. For amplifications where the final product was expected to be less than 2 kb, *pwo* DNA polymerase (Roche) was preferentially used. However as *Taq* DNA polymerase (Roche) is considered more robust than the *pwo* DNA polymerase, *Taq* DNA polymerase is was often used, especially with nested reactions.

For the 1985 samples each gene region was amplified to see if any insertions or deletions were present. All of the gene amplification products were used for sequencing except for the X-gene product, which overlaps with the precore/core and polymerase products.

3.3 Primer Design

The wild type hepatitis B virus genome is not circular in the virion but rather contains a nick at base pair 1850. Consequently, for complete amplification of the viral genome primers need to bind on either side of the nick. The genomic primers used in the project were based on Gunther's *et al.* (1995), refer to Table 3.1, primers HB1877F and HB1839R. The 3' ends of the primers are matched to HBV but contain SsfI and SapI enzymatic splice sites at their 5' end in case cloning is necessary.

To sequence the complete HBV genome, the 3.2kb genome was divided into five overlapping fragments of about 650 bp each. Many of the primers used in this work had been previously designed (Harrison, 1999). However new primers were required (refer to Table 1.1, those marked with * were designed for this project). New priming sites were chosen using an alignment of 253 genomes in Se-Al (Rambaut, 1995) which included non-human primate samples. Homogenous sites were identified in regions where primers were required. In the case of any heterogenous sites C and D genotypes were preferentially used as these are the only genotypes thus far found in New Zealand Maori. The primers were designed in Oligo 4 using these 8 basic rules:

- Complete conservation for 5 bp from 3' end, preferred 8 bp.
- At least 18 base pairs long with a melting point of at least 60°C
- No more than 3 degenerate nucleotide positions
- Base T will bind anything except itself (try to avoid at the 3' end)
- G, C anchor at 5' end (more G,Cs than A,Ts)
- 3' pentamers with a free energy (ΔG) value of no less than –9kcal/mol.
- Primer pairs no more than 6°C difference in Tm (prefer 2°C)
- Hairpin structures were no longer than 4bp with a ΔG value of no greater than -8 kcal/mol.

Table 3.1 HBV primers that were designed and used in this project and their purpose.

Primer	Primer purpose	Sequence	T_m °C
HB0050R	amp + seq	TGGAGCCACC AGCAGGRAAR TA	66
HB0400F	amp + seq	TTCCTCTKCA TCCTGCTGCT ATG	68
HB0547R	sequencing	CAGCAACADG AGGGAWRCWA TGAG	72
HB0680R *	amp + seq	CCACTGAACA AATKGCRCTA GTAAACTGA	76
HB0780F	sequencing	TGAGTCCCTT TWTRCCKCTR TT	54
HB0997R	geneamp + seq	CAGCAAANCC YAAAAGWCCC ACAA	68
HB1065F	amp + seq	CCCTTTTACR CAATGTGGNT ATC	66
HB1285R	amp + seq	GGAGTTCCGC AGTATGGATC G	66
HB1324F 2	Gene amp	CGATCCATAC TGCGGAACTC	62
HB1655F	amp + seq	CTTGGACTYY CWGSRATGTC	56
HB1720F *	Gene amp + seq	CCTTGAGGCM TACTTCAAAC AC	64
HB1774R 2	Gene amp + seq	AATTTATGCC TACAGCCTCC TA	62
HB1839R *	Amplification	GCTTGAGCTC TTCAAAAAGT TGCATGGTGC TGG	58
HB1849R *	amp + seq	TTGAGCTCTT CAAAAAGTTG C	58
HB1859R 1	Gene amp	AGCTTGGAGG CKTGAAMAGT	58
HB1877F *	Amplification	GCTTGAGCTC TTCTTTTTCA CCTCTGCCTA ATCA	58
HB1865F *	amp + seq	TTGAGCTCTT CTTTTTCACC	56
HB2057F	amp + seq	ACWGCACTCA GGCAAGMWRT T	60
HB2284F *	Gene amp + seq	GGTVTCTTTT GGAGTGTGGA TTC	66
HB2304R	amp + seq	GTKGATARGA TAGGGGCATT TG	60
HB2446F	amp + seq	ATGTTAGTAY YCCTTGGACT C	54
HB2487R *	Gene amp	GTAGAAGAAT AAAGCCCMGT AA	60
HB2667R 2	amp + seq	TTWGGATAAA ACCTAGCAGG CAT	64
HB2799F	Gene amp	AGYGCHTCAT TYTGYGGGTC AC	64
HB3121F *	amp + seq	CCKCCTCCTG CYTCCACCAA TC	70

*primers designed by me. Amp = primer used for amplifying product. Seq = sequencing primer.

3.4 Confirmation of PCR product

Each PCR reaction, including all controls, was visually checked for product. Ten percent of the reaction product was loaded onto 1% (w/v) TAE agarose gel and electrophoresed at 100V for 40 to 60 minutes depending on expected product size. Invitrogen 1 kb plus ladder™ and low DNA mass™ ladder were run beside the PCR product to check the that the product was the correct size and to estimate the DNA concentration. The gels were stained in ethidium bromide and photographed over ultra violet light for DNA visualisation.

3.5 PCR Template Purification

Due to their interference with sequencing reaction, it was necessary to remove all substrates from PCR product. The following two methods were used, subject to requirements and availability.

3.5.1 Shrimp Alkaline Phosphatase /Exonulcease I digest

Two units of Shrimp Alkaline Phosphatase (SAP) and 10 units of exonuclease I (Exo I) were added to the remaining 18 µl of short range reaction or 48 µl of long range reaction of PCR reaction mix to a total volume of 21µl or 51 µl, respectively. These reactions were incubated in the Thermal Cycler at 37°C for 30 minutes then 80 °C for 15 minutes. Exo I degrades single stranded oligomers and SAP removes phosphate groups. The end result is removal of all single stranded products such that they do not interfere with sequencing reactions. This procedure is used when the PCR product is a strong single band with no smearing and no primer dimers.

3.5.2. Rapid Gel Extraction

As per Concert™ Rapid Gel Extraction System protocol. The remainder of the PCR product was run on 1% (w/v) TAE agarose at 100mv until a clear separation of bands was achieved. The required fragment was excised from the agarose with a fresh sterile scalpel blade and weighed. The excised fragment was then incubated at 50°C in a 1.5 ml microcentrifuge tube with Gel Solubilising Buffer at a ratio of 30µl buffer to 10mg of gel, until the agarose was completely dissolved.

3.6 DNA Quantification

Automatic DNA sequencing reactions require the DNA template concentration to be within a specified range, which varies depending on the length of the DNA fragment. Therefore it is necessary to quantify the PCR product. This is done by running a known

volume of the template against 10 µl the Invitrogen low DNA Mass ladder™. Each band of the ladder is of a specified length and concentration. Comparison of the intensity of the band of PCR product to the control (ladder) enables quantification of the template.

3.7 Cloning

There were some PCR fragments of the viral samples that would not sequence directly, because of deletions or heterogenic sequence. Consequently cloning was employed to give sequence. The protocol used was as outlined in the Promega pGEM®-T and pGEM®-T Easy vector systems technical manual.

3.7.1 Ligation

The cleaned and quantified PCR fragments were ligated using the Promega pGEM®-T Easy vector system, in Life Technologies MAX Efficiency DH5α™ competent cells. A 2:1 insert vector molar ratio in 10 µl ligations was set up containing 1x T4 buffer, 50ng of Promega pGEM®-T vector, 2 units of T4 ligase, an appropriate amount of PCR product to give the 2:1 molar ratio required, and milliQ ddH_2O to give to 10 µl total volume. To allow complete ligation, the reaction was stored at 4°C overnight.

3.7.2 Transformation

A 30 µl aliquot of MAX Efficiency DH5α™ competent cells (30 µl) were used in the transformations. The entire ligation was added to the cells, and maintained on ice for 30 minutes. The cells were heat shocked at 37°C for 20 seconds and then placed back on ice for another two minutes. An aliquot of 200 µl of LB broth (1% (w/v) Tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, 1.5% (w/v) Agar) was then added, and the cells were gently shaken at 37°C for 1 hour. The cells were diluted as necessary and 100 µl was spread onto LB plates with 100µg/µl Ampicillin. The plates were incubated at 37°C overnight. *Lac* operon blue/white colony selection was used to identify transformants.

3.7.3 Recombinant DNA Extraction

Possible positives were picked and incubated overnight in 5 ml LB broth with 100 µg/ml ampicillin on a shaker. A 1.5 µl aliquot of each culture was centrifuged at 14,000 rpm for 1-2 minutes then resuspended in 350 µl of STET buffer (0.1M NaCl, 10mM Tris-HCl, 1mM EDTA, 5% Triton x-100). To this, 25 µl of lysozyme solution (10mg/µl in 10mM Tris, pH8) was added and the cells were heat shocked in boiling water for 40 seconds. The lysed cells were then centrifuged for 10 minutes at 14,000 rpm, and the debris was aseptically plucked out of the supernatant. The DNA was then precipitated for 15 minutes at room temperature in 40 µl 3M NaOAc (pH 5.2) and 420 µl isopropanol, then centrifuged for 15 minutes at 14,000 rpm to complete the precipitation. The DNA pellet was washed in 2.5 volumes of 70% ethanol, air dried and re-suspended in 25-50 µl of milliQ ddH_2O depending on the size of the pellet.

The presence of a PCR insert in the extracted plasmid DNA was determined by either an enzymatic digestion of a sample of the DNA or a direct PCR. For the enzymatic digest, 10 units of *Eco*R1, 1x digestion buffer, 5 µl of DNA from the extracted DNA sample were added together, and made up to a total volume of 25 µl with milliQ ddH_2O and incubated at 37°C for 1 hour. In the last 2 minutes of digestion, 2 µl of RNAse solution (2mg/µl milliQ H_2O) in was added to break down any RNA which interferes with gel staining. The digested product was then quantified on 1% (w/v) agarose gels, to check both for the presence of insert and to quantify the DNA extract.

3.8 Sequencing

Automated single stranded sequencing with ABI Prism® BigDye™ Terminator was the sequencing method chosen. The amount of DNA used in a sequencing reaction was estimated as the size of the product divided by 20, for example, for an 800 nt product the amount of DNA used was 40ng. The appropriate amount of DNA template was added to 4 µl of sequencing reagent and 4 µl of 5x sequencing buffer (400mM Tris-Cl, pH9.0 and

10mM $MgCl_2$) and 3.2 µl of primer (1pmol/µl) in a total volume of 20 µl made up with milliQ ddH_2O. Optimal sequencing reaction required the thermal cycle to undergo specific temperature graduations, thus the protocol used was as follows:

Step 1: 96°C for 10 seconds
Step 2: 50-55°C for 5 seconds
Step 3: 60°C for 4 minutes
Step 4: repeat from step 1 to step 3 for 24 cycles
Step 5: rapid ramp down to 4°C and hold.

Ramp speed is set as 1°C per second. Step 2 is variable because some primers anneal better at higher temperatures, though most of the annealing temperatures were 50°C.

3.8.1 Precipitation

The samples were then precipitated and dried using either a standard ethanol precipitation, carried out at room temperature or a standard isopropanol method.

3.8.1.1 Ethanol method

2.5 volumes of 95% ethanol, and 0.1 volume of sodium acetate (pH 4.7) were added into the sequence reaction, which was then precipitated for 15 minutes at room temperature. The precipitate was then centrifuged at 14,000 rpm for 30 minutes. The supernatant was removed and the pellet was washed with 200 µl of 70% (v/v) ethanol. After a brief vortex the precipitate was re-centrifuged at 14,000 rpm for 15 minutes. The 70% (v/v) ethanol was removed and the precipitate was left to dry in the dark.

3.8.1.2 Isopropanol method

For a 20µl DNA sequencing reaction, 80µl of 75% (v/v) isopropanol were added to each sample and the samples were precipitated for 20 minutes at room temperature in the dark,

followed by a spin at 14,000 rpm for 20 minutes. The supernatant was then removed and the pellet was washed with 250 µl of 75% (v/v) isopropanol. The precipitate was immediately re-centrifuged at 14,000 rpm for 5 minutes. The 75% (v/v) isopropanol was removed and the precipitate was left to dry in the dark.

The dried samples were then submitted to MuSeq, Massey University's sequencing facility, Palmerston North, for analysis on their ABI Prism™ 377 DNA Sequencer.

3.9 Sequence Analysis

The resulting electropherogram of the sequence data was processed through ABI Prism™ Sequencing Analysis Software™ using the BigDye™ Terminator mobility files. This software enabled visualisation of the peaks associated with the base calls. The entire genomes of individual samples were then imported into Sequencher 4.1 software and edited to create a consensus sequence.

The HBV genomes sequenced for this thesis were manually aligned in Se-Al (Rambaut, 1995) with 78 entire HBV-C genomes and 45 HBV-D genomes downloaded from GenBank (www.ncbi.nlm.nih.gov/). The consensus for each genotype was defined using a 70% cut off threshold, that is a base that was present at a frequency of 70% or greater was considered the 'consensus base'. The combined alignment was then exported into PAUP 4.0 (Swofford, 2000) for analysis.

3.10 HLA typing

Some of the host DNA extracted form the new Zealand samples was sent to the Transplantation Immunology group, Oxford Transplant Centre, Churchill Hospital, England, for HLA typing.

Chapter 4 Results and Discussion

The data will be described in two sections, starting with the analysis of the genomic sequences obtained from the HBV positive serum samples from Kawerau, their features and phylogenetic analysis. This is followed by analysis of the HLA typing of the individuals. Discussion of the results occurs throughout.

4.1 HBV Genome sequences

4.1.1 HBV sequencing results

50 HBV positive serum samples from 1998 and 1985 were provided of which 45 were amplified and continued with. Five of the 1985 samples were not available when the bulk of the work was being carried out, thus results from here will only deal with the remaining 45. Table 5.1 gives an overview of the clinical groups in the 1998 samples. Based on serological tests these samples fall into three clinical groupings. The first group have abnormal liver function and are positive for the HBeAg, and henceforth will be denoted as ++. Because the HBeAg is indicative of HBV expression this clinical group, was expected to have a higher viral titre. The second clinical group have abnormal liver function but are negative for the HBeAg +−. The last group have both normal liver function and are negative for HBeAg −−. The last two clinical groups (+− and −−) are negative for HBeAg, and have a lower viral titre compared to the first clinical group ++.

4.1.1.1 1998 samples

The first result is that of the 45 samples, 80% were genotype D and 20% were genotype C. Genotype A, the most frequent northern European strain, was not detected. All clinical types (++, +−, −−) were represented in the 20% of samples that are Genotype C, thus there is no evidence in this small sample for any effect of Genotype C on clinical type (Table 4.1).

Table 4.1 Clinical groups in the 1998 samples, by viral genotype.

Clinical groups	Genotype C (%)		Genotype D (%)		Total, clinical types
++	3	(23.1)	10	(76.9)	13
+-	3	(23.1)	10	(76.9)	13
--	3	(15.8)	16	(84.2)	19
Total, genotypes	9		36		45

++ HBeAg positive and abnormal for liver function tests (LFT);
+- HBeAg negative and abnormal for LFT;
-- HBeAg negative and normal for LFT.

From the 1998 samples, only the ++ clinical group was amplified and sequenced in the time available. The lower viral titre of the other clinical groups meant that more time would be required to get amplification products. Table 4.2 gives a summary of the genomes sequenced.

Table 4.2 Summary of sequencing results for the 1998 samples, ++ clinical type.

1998 genotype	No. of samples completely sequenced	(%)	No. of samples more than 75% sequenced	(%)	Total no. of samples (%)	
HBV- C	2	(25)	1	(20)	3	(23.1)
HBV- D	6	(75)	4	(80)	10	(76.9)
Total	8		5		13	

Eight of the 13 samples were successfully sequenced, the other 5 had heterogeneic sequence in the X protein/core/pregenomic promoter region (nt 1700-1800). Cloning was carried out on the PCR amplicons for these regions with mixed results, table 4.3 gives a summary.

Table 4.3 Cloned regions from 1998 samples.

Sample (1998)	Region cloned	Reason for not sequencing	No. of clone attempts	Cloned		Sequence results			
						Forward		Reverse	
				cl-1	cl-2	cl-1	cl-2	cl-1	cl-2
AN12	1630-2060	Not known	2	×	×	-	-	-	-
MO59	1670-1950	Heterogeneic sequence	2	×	✓	-	✓	-	✓
PA31	1600-2100	Not known	2	×	×	-	-	-	-
SA18	1600-2100	Not known	2	×	×	-	-	-	-
TA236	1600-2100	Not known	1	×	-	-	-	-	-
AN12	2060-2500	Heterogeneic sequence	1	✓	×	✓	-	✓	-
SA18	1150-1800	Deletion (8nt)	2	×	✓	-	✓	-	✓
WA13	2600-1-50	Deletions (54, 36, 11nt)	1	×	-	-	-	-	-

NB ✓= successful; × = not successful; - = not applicable; cl = clone.

4.1.1.2 1985 samples

I worked on all 45 of the 1985 samples. Moreover the gene regions were also amplified for identification of potential genomic insertions and deletions. Multiple bands were identified in most samples for at least one gene amplification (Figure 4.1). This demonstrated that even at this earlier stage of infection the viral genomes existed in a quasispecies state.

Of the 1985 samples, all except one (TA236), were successfully amplified for at least one gene region, most samples successfully amplified for all gene regions. TA236 did not amplify at all, either it had a lower viral titre, or a bad extraction, or both. The results of the PCR amplification of the viral genomes are summarised in Table 4.4. Amplification of the polymerase gene was split into two smaller amplifications (Pol-1 and Pol-2) that were bisected in the middle by the Surface gene region; the reason for creating these smaller products was to get better sequencing results.

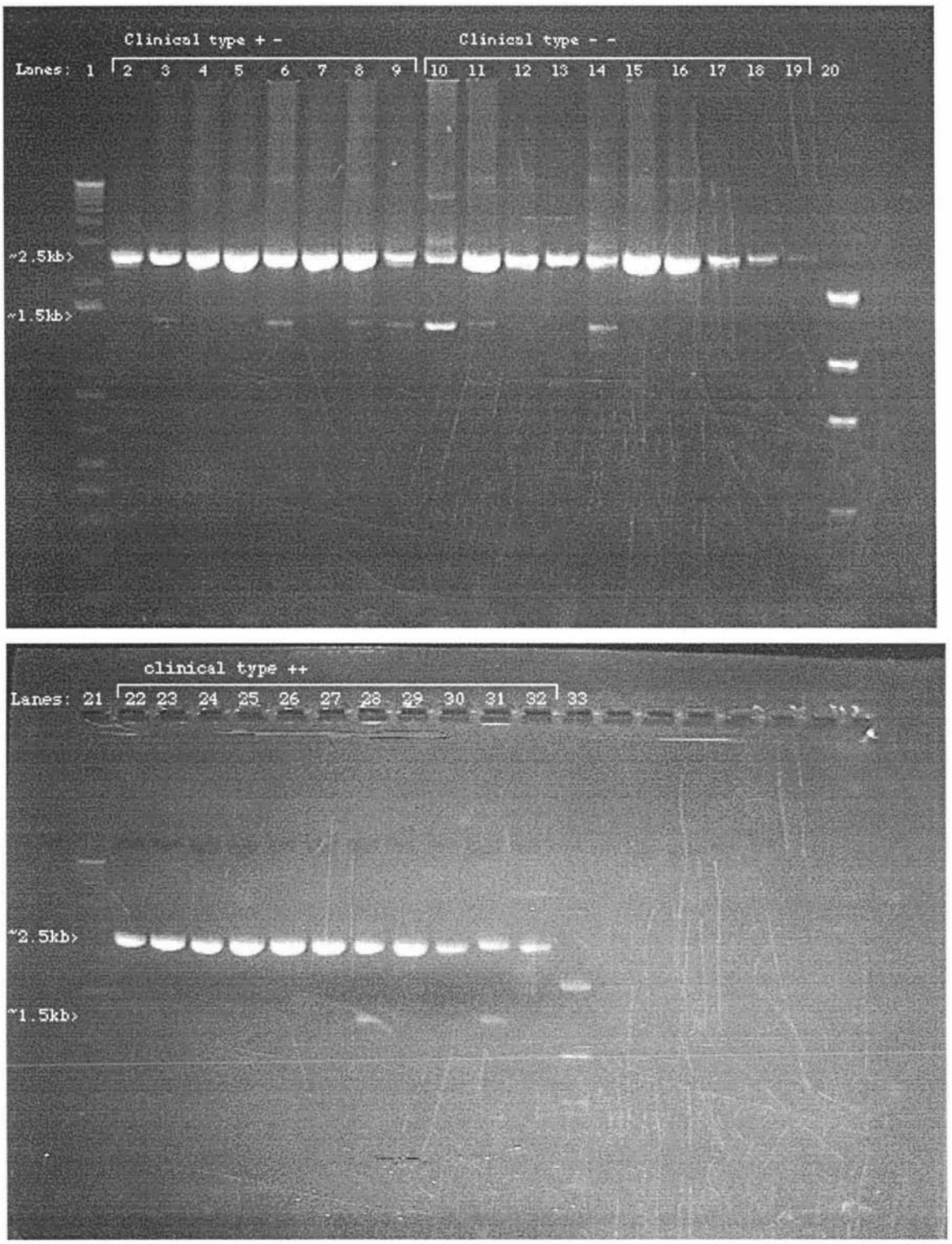

Figure 4.1 Gels shows samples from clinical types +- and -- and ++ amplified for the Polymerase gene region (about 2.5kb in size). Samples in lanes 3, 6, 8, 9, 10, 11, 12, 14, 28 and 31 all have an extra band at around 1.5kb in size. Samples in lanes 9, 18, 19, 31 and 32 are genotype C, all other samples are genotype D.

Table 4.4 Gene regions that were amplified of the 1985 samples.

Sample		Amplified Areas									
		X gene 1316-1840		PreC/C 1816-2454		S gene 2850-1-837		Pol 2309-1-1625		Pol-1* 2284-1-50	Pol-2* 780-1174
		amp	indel	amp	Indel	amp	Indel	amp	Indel	amp	amp
AN12	++	✓		✓		✓		✓	✓✓		
BU12	++	✓		✓	✓✓	✓		✓	✓✓		
CH02	++	✓		✓	✓✓	✓		✓	✓✓		
HA53	++	✓		✓	✓✓	✓		✓	✓✓		✓
MO59	++	✓		✓	✓✓	✓		✓			
PA31	++	✓		✓		✓		✓		✓	
PA58	++	✓		✓	✓✓	✓		✓	✓✓	✓	✓
SA18	++	✓		✓		✓	✓✓	✓	✓✓	✓	✓
TA40	++	✓		✓	✓✓	✓	✓✓	✓		✓	✓
TA236	++	×		×		×		×		×	×
TE55	++	✓		✓	✓✓	✓	✓✓	✓		✓	✓
WA13	++	✓		✓		✓	✓✓	✓		✓	✓
AN10	+−	✓		✓		✓		✓	✓✓	✓	✓
AN19	+−	✓		✓		✓		✓	✓✓		✓
BE03	+−	×		✓	✓✓	✓		×		✓	✓
BO13	+−	×		✓		✓		×		✓	✓
HU27	+−	✓		✓	✓✓	✓		✓	✓✓	✓	✓
KA40	+−	✓	✓✓	✓		✓		✓			
MO70	+−	✓		✓	✓✓	✓		✓		✓	
ORO3	+−	✓		×		✓		×		✓	w
PE04	+−	×		×		×		×		✓	w
PE09	+−	✓		✓	✓✓	✓	✓✓	✓	✓✓	✓	✓
RA11	+−	✓		✓		✓	✓✓	✓		✓	✓
TA120	+−	✓		✓	✓✓	✓	✓✓	✓	✓✓	✓	✓
TA239	+−	×		✓		✓		×		✓	w
ADO3	−−	✓		✓	✓✓	✓		✓	✓✓	✓	✓
AD05	−−	✓		✓		✓		✓	✓✓	✓	✓
AN20	−−	×		✓		✓		×		✓	✓
BU32	−−	×		✓	✓✓	×		×		✓	w
CH04	−−	×		w		✓		×		✓	w
CI11	−−	×		✓		✓		✓	✓✓	✓	✓
CO04	−−	×		w		×		×		✓	w
GE06	−−	×		✓		×		×		✓	w
GE09	−−	×		✓		✓		w		✓	w
HI43	−−	✓	✓✓	✓	✓✓	✓		✓	✓✓	✓	✓
HU11	−−	✓		✓	✓✓	✓		✓	✓✓	✓	✓
HU29	−−	✓		✓	✓✓	✓		✓	✓✓	✓	✓
MI26	−−	✓	✓✓	✓		✓		✓		✓	
MI27	−−	×		✓		✓		×		✓	w
PA81	−−	✓		✓		✓		✓		✓	✓
RA07	−−	×		✓	✓✓	×		×		✓	w
TA229	−−	✓		✓	✓✓	✓	✓✓	✓		✓	✓
TU25	−−	✓		✓		✓	✓✓	✓		✓	✓
WO06	−−	×		✓		✓		×		✓	w

Amp refers to PCR amplified. In 'amp' columns, ✓= successful; × = not successful; w = weak (less than 5ng/µl); blank cell = not applicable. Indel refers to insertion or deletion seen from banding patterns on the gel, ✓✓ = multiple bands, blank cell = either single band or no bands.
* Pol-1 and Pol-2 PCR amplicons do not overlap.

Routine PCR amplification and sequencing as described in the Materials and Methods chapter had a various success rate for the 1985 samples. The sequencing results for the 1985 samples are summarised in Table 4.5.

Table 4.5 Summary of sequencing results for the 1985 samples, breaking down the results into genotypes and clinical types.

1985 genotype & clinical type	No. of samples completely sequ'ed (%)	No. of samples more than 70% sequ'ed (%)	No. of samples more than 50% sequ'ed (%)	No. of samples less than 50% sequ'ed (%)	Total no. of samples
HBV-C	0 (0)	2 (25)	1 (14.3)	6 (33.3)	9
HBV-D	12 (100)	6 (75)	6 (85.7)	12 (66.7)	36
Total (genotype)	12	8	7	18	45
++	6 (50)	2 (25)	2 (28.6)	3 (16.7)	13
+−	3 (25)	3 (37.5)	2 (28.6)	5 (27.8)	13
−−	3 (25)	3 (37.5)	3 (42.9)	10 (55.6)	19
Total (clinical)	12	8	7	18	45

Genotype D virions sequenced more readily than genotype C virions. Only 3 (33.3%) of the 9 genotype C samples have been more than 50% sequenced, whereas 24 (66.6%) of the 36 of the genotype D samples are more than 50% sequenced, this is significant ($p<0.05$). Why genotype D is sequenced more readily over genotype C was never completely determined, possibly viruses that are genotype D have a higher viral titre than genotype C. It would be interesting in the future to test this possibility with real time PCR, though differences in the sample extractions would limit this. Several lines of trouble shooting were carried out for any regions that gave unsatisfactory sequencing results. Variations of extension times, primer annealing temperatures, and primer combinations, as well as new protocols for DNA precipitation, were all tried. Overall more than 120,000 base pairs of HBV DNA was sequenced during this project!

4.1.1.3 1985 and 1998 sample comparison

All the virus genomes (1985 and 1998 samples) that were over 70% sequenced were aligned and analysed. All results from here on pertain only to these samples. Table 4.6 lists those samples and gives their clinical types and genotypes.

Table 4.6 Samples which were sequenced over 70%, and were analysed further.

Sample	Year	Clinical type 1998	Genotype
AN12	1985	++	D
BU12	1985	++	D
CH02	1985	++	D
HA53	1985	++	D
MO59	1985	++	D
PA31	1985	++	D
PA58	1985	++	C
SA18	1985	++	D
TA40	1985	++	D
TE55	1985	++	C
WA13	1985	++	D
AN10	1985	+-	D
KA40	1985	+-	D
TA120	1985	+-	D
AD05	1985	--	D
CI11	1985	--	D
HI43	1985	--	D
AN12	1998	++	D
BU12	1998	++	D
CH02	1998	++	D
HA53	1998	++	D
MO59	1998	++	D
PA31	1998	++	D
PA58	1998	++	C
RA37	1998	++	D
SA18	1998	++	D
TA40	1998	++	D
TA236	1998	++	C
TE55	1998	++	C
WA13	1998	++	D

These samples were compared to a consensus of their corresponding genotype sequence and checked for mutations. The consensus sequence for each genotype was constructed by downloading and aligning 253 available HBV genomes from GenBank (www.ncbi.nlm.nih.gov/). Genotypes D and C sequences were identified and separated into HBV-D and HBV-C files. The consensus was identified using a 70% cut off threshold, that is, a base that was present at a frequency of 70% or greater was considered the 'consensus base'. Mutations were checked for in areas of significance, such as epitopes, promoter regions, enhancer regions, functional motifs (see appendix A for summary tables). All 1998 samples belong to the ++ clinical group.

4.1.2 X-gene (nt 1316-1840)

The regions of interest that contain mutations for the X gene region are:

1. A hypervariable region (aa 36-47; nt 1463-1516).
2. The region of X Associated Protein 1 (XAP-1) binding (aa 55-101; nt 1538-1677). XAP-1 is a DNA repair protein.
3. The core promoter stretches from nt 1591-1851 and contains all of the following areas of interest.
    4. Direct repeat I (aa 73-76; nt 1592-1602).
    5. Direct repeat II (aa 152-155; nt 1828-1838).
    6. The enhancer II region (aa 105-144; nt 1687-1805); which contains both:
        7. The 1762-1764 double mutant region (aa 130 and 131); and
        8. the functional domain for translation (aa 110-139, nt 1702-1792).

The mutations seen in these areas are mostly single nucleotide mutations. The most variable of these regions is of course the hypervariable region. Table 4.7 gives a summary of the X gene results table.

This summary shows that the 1998 samples are more variable than the 1985 samples (58.6% and 41.4% respectively). And that the samples that were in the ++ clinical type in 1998 are more variable than the others.

Table 4.7 The number of samples with mutation(s) for the X-gene region, showing differences between the clinical types.

X-gene region	1	2	3	4	5	6	7	8
aa	36-47	55-101	73-158	73-76	152-155	105-144	130-131	110-139
nt	1463-1516	1538-1677	1591-1851	1592-1602	1828-1838	1687-1805	1762-1764	1702-1792
++ 1985	4	0	1	1	1	0	0	0
+−	2	0	0	0	0	0	0	0
−−	0	1	0	0	0	1	1	0
++ 1998	6	0	1	1	0	1	5	3

The mutation within Direct repeat I (aa 73-76, nt 1592-1602) that causes a peptide change from phenylalanine (aa 73) to valine occurs in HA53_1985 and HA53_1998 (both ++). A known mutation at this site, aa 73 phenylalanine to leucine, is often effectively silent (Uchida et al., 1997). The change to valine may also be silent as valine is also a hydrophobic aa, like leucine. However, phenylalanine is considered more hydrophobic than both valine and leucine due to its aromatic ring structure. The two repeats, DR1 and DR2 are implicated in the HBV DNA synthesis (Uchida et al., 1997), so a nucleotide change within either of these regions would probably have an effect on HBV DNA synthesis.

The double mutation at nucleotides 1762 and 1764 (aa 130 and 131) removes a nuclear binding receptor site in the core promoter, suppressing both precore and core RNA transcription. A HNF1 transcriptional site is created by the double mutation, restoring the core RNA level (Li et al., 1999). This double mutation has previously been associated with HBeAg-negative patients with severe liver damage (Hannoun et al., 2000). The nucleotide changes result in the following amino acid changes: lysine to methionine (aa 130) and valine to isoleucine (aa 131). Lysine is a basic amino acid with an overall positive charge, whereas methionine is an apolar neutral amino acid which is hydrophobic. Valine is an apolar neutral amino acid, which is hydrophobic, whereas isoleucine is a polar neutral amino acid that is not hydrophobic. HI43_1985 (of the -- clinical group), HA53_1998, and RA37_1998 each have this double mutation. SA18_1998 has a triple mutation at this site (AAG GTC to ATT ATC) resulting in lysine (130) changing to isoleucine instead of methionine.

Both the SA18_1985 and SA18_1998 samples have an 8 base-pair deletion from 1765-1773, which is in the middle of enhancer II and the functional domain for translation. An 8 nt deletion will of course make that copy non-functional, but there will still be a minority that have the full length transcript. The deletion includes the amino acid 132 (the highly hydrophobic phenylalanine), which (along with aa 69 and 61) is essential for X protein activity (Becker et al., 1998). The TE55_1998 sample has a similar deletion of 9 base pairs from 1760-1769, this deletion includes the double mutation at nucleotides 1762 and

1764, and the essential amino acid 132. X protein activity is either completely abolished or severely retarded in all these virions.

WA13_1998 has either a G or a T at 1758, giving either an arginine or an isoleucine at aa 128. Arginine is basic with an overall positive charge, and isoleucine is neutral amino acid. There are two more nucleotide changes within the enhancer II element, these are T to A at 1797 in the -- sample AD05_1985, and C to G at 1801 in the ++ sample WA13_1998. Both these nucleotide changes are situated at the 3' end of enhancer II element. The affect of these changes would have to be determined by further experiments into protein function.

There are two samples, both ++, with nucleotide changes within the core promoter which are not associated with other regions, these are an A to C at 1678 in TE55_1998 and A to C at nt 1821 in PA31_1985. The change at nt 1678 falls in between the region required to react with XAP-1 and the enhancer II element and lies within the 5' half of the core promoter. The change at nt 1821 is just before the Direct Repeat II, at the 3' end of the core promoter. And finally the ++ SA18_1985 has either a C or a T at nt 1833, which is within the Direct Repeat II. As DR2 is implicated in HBV DNA synthesis (Uchida et al., 1997), this nucleotide change would probably reduce HBV DNA synthesis.

4.1.3 Precore/Core region (nt 1816/1903-2454)

The precore/core (preC/C) region spans nucleotides 1816-2454, the core gene starts at nucleotide 1903. The regions of interest that contain mutations (see Table 4.8) within the precore/core regions are:

1. The functionally important region of encapsidation sequence (nt 1852-1930). This includes the 1896 stop mutation.
2. CTL epitope (aa 18-27 of core; nt 1952-1981).
3. T-cell epitope (aa 50-69 of core; nt 2052-2109).

4. HBcAg and HBeAg B cell epitopes; (aa 74-83; nt 2122-2151), (aa 76-89; nt 2128-2169), (aa 107-118; nt 2221-2256), (aa 128-135; nt 2284-2307), (aa 130-138; nt 2290-2316).
5. N-terminal functional domain (aa 80-130; nt 2140-2292).
6. CTL epitope (aa 141-151; nt 2323-2355).

Table 4.8 The number of samples with mutations in the function areas of the PreC/C gene region, showing differences between the clinical types.

PreC/C gene region	1	2	3	4					5	6
aa	-	18-27	50-69	74-83	76-89	107-118	128-135	130-138	80-130	141-151
nt	1852-1930	1952-1981	2052-2109	2122-2151	2128-2169	2221-2256	2284-2307	2290-2316	2140-2292	2323-2355
++ 1985	0	0	0	0	1	0	2	0	0	2
+-	0	0	1	0	0	0	0	0	1	0
--	1	0	0	1	0	1	1	0	1	1
++ 1998	3	1	5	3	0	0	1	1	4	1
Total	4	1	6	4	1	1	4	1	6	4

Again there are more samples with mutations in the 1998 samples than in the 1985 samples, 59.4% and 40.6% respectively. The four mutations within the functionally important region for the encapsidation signal (nt 1852-1930) all fall within the precore gene. SA18_1998 at nucleotide 1874 has either an A or a C, which gives either a threonine or proline at amino acid position 21 of the precore protein. At the secondary structure level of the DNA, nucleotide 1874 is in the balloon part of a stem structure, therefore this mutation is not as important compared to a change in the stem structure.

The -- HI43_1985 and the ++ RA37_1998 both have the well documented 1896 G to A mutation, (see Figure 2.1), which changes a tryptophan to a stop codon at amino acid 28 of the precore protein. This mutation functions as an immune response escape mutant, because it prevents the expression of the HBeAg on the surface of HBV infected hepatocytes, and therefore protects these cells from being eliminated by cytotoxic T cells specific for the HBeAg (Arauz-Ruiz et al., 1997). Virions with this mutation are HBeAg negative, yet RA37_1998 is in the ++ clinical group, which is positive for HBeAg. This mutation in RA37_1998 must be relatively recent, or there is a mixed viral population -

some with the mutation, some without. It would be interesting to determine the strength of the serological response.

WA13_1998 is heterogeneic, with both A and G at nucleotide 1899. This nucleotide change (G to A) is associated with the secondary stem structure of DNA, it increases the stability of the stem by providing an additional paired site (nucleotide 1899 binds with 1855, which is a T) (see Figure 2.1). Nucleotide 1858 is opposite nucleotide 1896 in the stem structure, and is usually a T. An increase in the stability of the stem structure occurs with the G to A mutation at 1896. There can also be a T to C mutation at 1858 which prevents the 1896 mutation, as it destabilises the stem of the encapsidation signal. All of the samples which I analysed had a T at 1858, which is common in genotype D variants (Arauz-Ruiz et al., 1997).

All other mutations are within the core gene region, and the amino acid numbering pertains to the core gene region. The first is a C to T change at nucleotide 1971 in RA37_1998. This is within the CTL epitope (aa 18-27; nt 1952-1981), this mutation has no effect at the peptide level because amino acid 24 remains phenylalanine.

HBV core antigen variants accumulate mutations in their T and B cell epitopes, which successfully avoids detection by the immune response. The following mutations are all within epitope regions and thus will affect immunological responses.

- The first is an A to C change at nt 2073 in RA37_1998, resulting in an amino change from glutamine (aa 53) to histidine. Glutamine is a neutral amino acid, whereas histidine is basic with an overall positive charge.
- Also in RA37_1998 is a C to T nucleotide change at nt 2075, resulting in amino acid change from a neutral hydrophilic alanine (aa 54) to a neutral hydrophobic valine.
- SA18_1998 has a nucleotide change G to A at nt 2092, which results in the acidic, negatively charged glutamic acid (aa 64) changing to the basic, positively charged lysine.

- Lastly several samples (+− TA120_1985, and the ++ CH02_1998, MO59_1998, RA37_1998, WA13_1998) have the nucleotide change A to T at nt 2094, which results in glutamic acid (aa 64) changing to aspartic acid.

Many of the HBcAg and HBeAg B-cell epitopes overlap. The first two of these epitopes (aa 74-83 and 76-89) overlap, and so any mutations there affect both epitopes.

- HI43_1985, clinical type −−, has a change of G to C at nt 2131. This mutation results in a glutamic acid to glutamine amino acid change.
- Nucleotide changes A to G at nt 2140 in CH02_1998 and WA13_1998; and nucleotide changes T to C at nt 2141 in HI43_1985 and MO59_1998, cause amino acid changes isoleucine to valine and isoleucine to threonine at amino acid 80, respectively.
- A mutation change which only occurs in the second epitope (aa 76-89) is either an A or a G at nucleotide 2155 in SA18_1985, which gives either a valine or an isoleucine at aa 85.

The N-terminal functional domain overlaps with the HBcAg and HBeAg B-cell epitopes. There are some mutations though within the N-terminal functional domain which do not fall into any of the HBcAg and HBeAg B-cell epitopes regions.

- For instance, the A to C nucleotide change at 2176 in PA31_1998, which causes the amino acid change asparagine to histidine at amino acid 92.
- TA120_1985 (clinical type +−), WA13_1998, and CH02_1998 have the nucleotide changes A to G at nt 2179, which results in the amino acid change methionine to valine at amino acid 93.
- MO59_1998 has a nucleotide change from G to A at nt 2181, which causes an amino acid change from methionine (aa 93) to isoleucine.
- MO59_1998 has another change from G to C at nt 2183, causing the amino acid change glycine (aa 94) to alanine.
- In other studies amino acid 99 is a mutational 'hotspot', but only HI43_1985 has a nucleotide change here, A to G at nt 2199. The amino acid remains glutamine with this change.

HI43_1985 (clinical type --) has a double mutation at nucleotides 2240 and 2241, GA to CY. The amino acid differences are glutamic acid (aa 113) to either leucine or proline. This double mutation is within the third HBcAg and HBeAg B-cell epitope (aa 107-118).

The last two HBcAg and HBeAg B-cell epitopes (aa's 128-135 and 130-138, respectively) overlap with the 3' end of the N-terminal functional domain, and overlaps partly with the B-cell epitope (aa 120-130, nt 2260-2292) and a T-cell epitope (aa 117-131, nt 2251-2295). Amino acid 130 falls within the overlap of all these functional regions, and is a mutational hotspot. CH02_1985 (clinical type ++) has either a C or a T at nt 2290, resulting in either proline (130) or serine. HI43_1985 (clinical type --), and WA13_1998 both have the nucleotide change C to A at 2291, resulting in proline (130) to glutamine amino acid change. PA31_1985 (clinical type ++) has either a C or an A at nt 2291.

The last HBcAg and HBeAg B-cell epitope has a mutation that is unique to it, a C to A change at nt 2307 in PA31_1998, causing an amino acid change from proline (aa 135) to histidine. Most of the mutations within the B-cell epitopes are in the ++ clinical group, this is congruent with past research (Carmen et al., 1997).

Several samples have mutations in the HLA class I restricted CTL epitope (aa 141-151, nt 2323-2355).

- CH02_1985 has a C to T change at nt 2325, causing a serine (aa 141) to leucine amino acid change.
- PA31_1998 has an A to T change at nt 2327, resulting on a threonine (aa 142) to serine change.
- PA31_1998 has a double mutation (AC to TG at nt 2342 and nt 2343) causing amino acid 147 to change from threonine to cysteine.
- AN12_1985 (++) has a T to G change at nt 2346, which results in a valine (aa 148) to glycine change.
- The G to A change at nt 2348 in the -- HI43_1985 and the ++ PA31_1998 results in a valine (aa 149) to isoleucine change.

With all of these mutations, 5 of the 11 amino acids in this epitope have been modified. This CTL epitope is restricted to HLA class I allele families HLA-A*31 and HLA-A*68. These HLA-A alleles do not feature in any of the hosts with the viruses that have the mutations within this HLA epitope.

4.1.4 Surface Gene (2850-1-837)

The surface gene transcript spans nucleotides 2850-1-837, and codes for three in-frame proteins, which all end at the same termination site: the preSI (start site nt 2850) produces the Large protein; preSII (start site nt 3207) transcribes the Middle protein; and the S (start site nt 157) produces the Small protein. Figure 4.2 is a schematic representation of the Surface gene region.

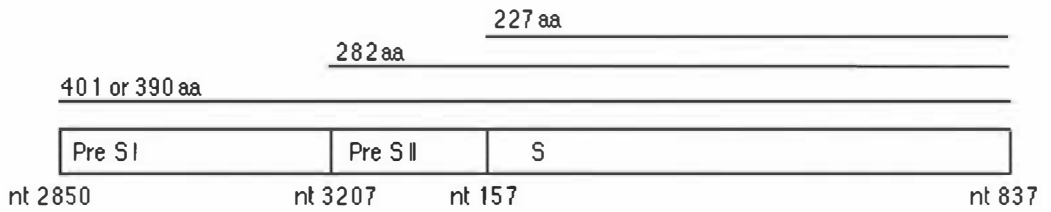

Figure 4.2: A schematic diagram of the Surface gene region. The Large protein is either 401 or 390 amino acids depending on the genotype.

PreSI contains the putative receptor binding region covering nucleotides 2910-2990, and within this region is the preSI epitope (aa 31-35; nt 2940-2954). Neither of these regions accumulated any mutations. In the preSI region there is a 33 nucleotide variable region (aa 2-12; nt 2853 to nt 2886) which is deleted in Genotype D and present in Genotype C. Other deletions in this area are found in SA18_1998 and WA13_1998, both of which belong to the ++ clinical type and are genotype D. These are:

- SA18_1998 has a 64 base pair deletion from nt 3029 to nt 3093 (aa 60 to 78).
- WA13_1998 has a 36 base pair deletion from nt 3046 to nt 3082 (aa 66 to 75); and an 11 base pair deletion from nt 3176 to nt 3187 (aa 106 to 110).

- Premature termination sites occur in two samples within the preSI region. PA58_1998 has a stop codon at nt 3006 (aa 52 of the preSI region). WA13_1998 has a stop codon at nt 3088 (aa 77 of the preSI region).

These early termination sites would not affect the expression of the downstream preSII or S genes, as they have a separate promoter to preSI gene. The Large protein would be affected because both termination sites occur at the 5' end, thus virion production would be reduced.

As the Large protein contains the preSII and S regions all mutations within these regions may affect the Large protein as well. The preSII region is relatively short (55 amino acids) and contains an asparagine at amino acid position 4 which is N-glycosylated at post-translation stages. However, there is a 54 base pair deletion from nucleotides 3 to 57 in WA13_1998. This deletion affects amino acids 4 to 21, but the asparagine still occurs, though it may no longer be in the correct position for N-glycosylation.

The S gene region (nt 157 to 837) is in all three proteins, and contains most of the regions of interest; those regions that contain mutations in these samples are:

1) Novel epitope for CTL responses (aa 5 to 12; nt 169-192).
2) Signal I (aa 8-22; nt 178-222) and
3) Signal II (aa 80-98; nt 394-450) are hydrophobic regions that are anchored in the Endoplasmic Reticulum (ER) membrane.
4) *d/y* determinant (aa 122; nt 520-522).
5) *a* determinant (aa 124-147; nt 526-597).
6) Epitope cluster (aa 154-159; nt 616-633).
7) CTL epitopes Env.362 (aa 362-371; nt 751-781) and Env.364 (aa 364-372; nt 757-784).

Most of the mutations within the S gene region are point mutations or degenerative nucleotide positions. Table 4.9 gives a summary of the S gene region.

Table 4.9 The samples with mutations in the functional areas of the Surface gene region, showing differences between the clinical types.

S gene region	1	2	3	4	5	6	7
aa	5 - 12	8-22	80-98	122	124-147	154-159	362-372
nt	169-192	178-222	394-450	520-522	526-597	616-633	751-784
++ 1985	0	0	1	0	0	0	0
+−	0	0	1	0	0	0	0
−−	0	0	0	1	0	1	1
++ 1998	1	1	2	0	3	0	1

NB this table is not including the preSI and preSII regions.

As expected, more 1998 samples have mutations than the 1985 samples, 8 and 5 respectively. For the whole of the Surface gene region, the difference between the 1998 and 1985 samples is greater, because all of the mutations within the preSI and preSII regions are all in 1998 samples.

CH02_1998 has a mutation within the CTL epitope (aa 5-12; nt 169-192) where there is a G to A nucleotide change, resulting in a glycine (aa 7) to glutamic acid change. This is a neutral amino acid to an acidic amino acid.

The two mutations which affect signal I (aa 8-22; nt 178-222) are both in TA236_1998, the first is an A to C change at 203, resulting in glutamine (aa 16) to proline amino acid change. Both are considered to be neutral amino acids, though proline may have different structural effects. The second mutation is a T to G at 209, resulting in valine (aa 18) to glycine amino acid change. Valine is hydrophobic whereas glycine is neutral. Signal I and signal II are both hydrophobic areas as they lie within the ER membrane.

Signal II (aa 80-98; nt 394-450) has more mutations than signal I.

- KA40_1985, a sample that falls into the +− clinical type, has a degenerative site at nt 416 where there is either a G or a T. Therefore amino acid 87 is either a hydrophobic leucine or a basic arginine, which has a positive charge, and is not hydrophobic.
- PA58_1985 and PA58_1998 both have changes at nucleotides 429 and 447. At 429 there is a C to T change, which is a silent mutation as amino acid 91 remains a leucine.

The mutation at 447 (T to G) is also a silent mutation, as amino acid also remains a leucine.

- PA58_1985 has an addition degenerative site at nt 430, where it is either an A or a C, which results in either an isoleucine or leucine at amino acid 92.
- TA236_1998 has a degenerative site at nt 448, where there is either a G or a C, resulting in either a leucine or a valine at amino acid 98. These are both hydrophobic amino acids, so there is no functional change.

Amino acid 122 (nt 520-522) is important for the *d/y* determinant. If there is a lysine here then *d* is specified, but if an arginine is at this position then *y* is specified. Genotype D variants will have a *y* specified, whereas genotype C variants will have the *d* specified. All the genotype D samples have an arginine at this position, and all the genotype C samples have a lysine at this position. HI43_1985 (clinical type --) has a silent point mutation a nt 522 (A to G) where the arginine remains arginine. Both lysine and arginine are basic amino acids with a positive charge.

The *a* determinant region (aa 124-147; nt 526-597) is also known as the dominant epitope cluster, and is within the Major Hydrophilic Region. WA13_1998 has an A to G change at nt 547, which results in the amino acid change threonine (aa 131) to alanine. Both threonine and alanine are neutral amino acids. Within the *a* determinant region is the *a* determinant loop (aa 139-147; nt 571-597). This loop is stabilised by a disulphide bridge that occurs between two cysteine residues (aa 139 and 147). Two samples have point mutations within this loop. In RA37_1998 a C to T change at nt 575 results in the change from threonine (aa 140) to isoleucine. HA53_1998 has a C to T change at nt 584, resulting in a change from serine to leucine at amino acid 143. This is a change from a neutral, slightly hydrophilic amino acid to a hydrophobic amino acid. The affect of this particular change is unknown.

The epitope cluster which spans amino acids 154 to 159 (nt 616-633) has a degenerative nucleotide in AD05_1985 (clinical type --). There is either an A or a G at nt 632, which is either a glycine (aa 159) or a glutamic acid.

The last 56 amino acids of the Surface region (nt 664-837) are very hydrophobic as the protein traverses through the ER membrane twice. The overlapping epitopes Env. 362 and Env. 364 (nt 751-784) are situated in this hydrophobic region. PA58_1985 (clinical type ++) has a degenerative base at 767, where there is either an A or a G, resulting in either a asparagine (aa 367) or a serine. Both these amino acids are neutral. AD05_1985 (clinical type --) also contains a degenerative base within these epitopes. There is either an A or a C at nt 780, this is a silent mutation as isoleucine at amino acid 371 remains isoleucine.

There are two point mutations that result in early termination sites within the Surface gene region. The first is in SA18_1998 at nucleotide 363, shortening the Large, Middle, and Small proteins by about 330 amino acids. The second is on WA13_1998 at nucleotide 803 which is not so drastic, as only 11 amino acids are lost from the N-terminal end of the proteins.

4.1.5 Polymerase

The polymerase protein consists of 3 functional domains and a non-essential spacer arranged in the following order: terminal protein (nt 2309-2839), spacer (nt 2840-1-131), reverse transcriptase/DNA polymerase (nt132-1163), and ribonuclease H (nt 1164- 1625). All four domains are expressed as a single translational unit.
The polymerase region completely overlaps with the Surface gene region, but is read in a different reading frame. The point mutations resulting in early termination sites seen in the Surface gene region are not early termination site mutations in the Polymerase reading frame. Having said this, there is one sample, CH02_1985 (clinical type ++), that has a point mutation at nt 2325 that results in an early termination site. This point mutation changes a Glutamine (aa 6) to a stop codon. The protein that results is a severely shortened protein, as this point mutation occurs at the 5' end of the terminal protein.

The 33 nucleotide (nt 2854-2887) deletion characteristic for genotype D variants affects both the surface and polymerase genes. This deletion occurs within the spacer domain of

the Polymerase gene. The spacer region can be deleted to a large extent without significant loss of endogenous polymerase activity (Radziwill, 1990). The preSI region overlaps with the spacer region, thus the deletions observed in WA13_1998 and SA18_1998 that are within the preSI region are also within the spacer region and have no affect here.

There are many point mutations within the polymerase gene region, in fact all of the samples have at least one point mutation in this region. But no mutations occur within any areas of biological, functional or structural importance, such as the YMMD motif. The YMMD motif is four amino acids (tyrosine – methionine – methionine – aspartic acid) that is situated at nucleotides 738-750 and is within the reverse transcriptase domain and is essential for activity. For the analysis of the Polymerase gene, the frequency of mutations for each of the Polymerase gene domains was calculated, see Table 4.10.

Table 4.10 Overall single nucleotide mutation rate for the different Polymerase domains.

Polymerase Domain	Terminal protein	(%)	Spacer region	(%)	Reverse transcripase (%)		Ribonuclease H	(%)
Total no. of Nucleotides	531		507		1032		462	
No, of point mutations	54	(10.2)	45	(8.9)	105	(10.2)	64	(13.9)

The overall mutation rates for each domain do not show much variance, so the number of silent to replacement mutations within each domain was compared. The ratio of silent to replacement change is a measure of functional constraint within a gene, or the degree to which a gene is conserved (Bowyer and Sim, 2000). For neutral changes the proportion of silent to replacement mutations is expected to be 0.43 (3:7) (Nei, 1987). Table 4.11 gives a summary of that analysis.

Table 4.11 Mutation frequencies for each domain in the Polymerase region, showing a break down of silent and non-silent mutations.

Genotype	Terminal protein nt 2309-2839			Spacer domain nt 2840-1-131			Reverse transcriptase nt 132-1163			Ribonuclease H nt 1164-1625		
	s	r	s/r	s	r	s/r	s	r	s/r	s	r	s/r
HBV-D	19	21	0.9	7	23	0.3	30	26	1.2	35	8	4.4
HBV-C	8	6	1.3	4	11	0.4	26	23	1.1	15	6	2.5
Total	27	27	1.0	11	34	0.3	56	49	1.1	50	14	3.6
Total Δ per nt, %	5.1	5.1		2.2	6.7		5.4	4.7		10.8	3.0	

Δ = changes; s = silent; r = replacement; s/r = silent/replacement ratio.

This data shows that the spacer domain has a lower silent to replacement mutation ratio than other domains, average s/r = 0.33 compared to average s/r ratios of 1.1 and 3.5. A functionally important gene will be well conserved with a silent to replacement ratio close to 1 or higher, since the main changes are silent (Bowyer and Sim, 2000). Ribonuclease H domain has the lowest number of replacement mutations compared to the other P gene domains. These results are a reflection of lack of functional constraints of the spacer region compared to Reverse transcriptase and Ribonuclease H regions.

Not all the samples had mutations in every P gene domain; Table 4.12 shows a summary of the number of samples mutated in each of the P gene domains.

Table 4.12 Number of samples with single nucleotide mutations in the different P gene domains, showing the silent and replacement mutations and the distribution of the mutations for each clinical group.

P gene regions	Terminal Protein			Spacer domain			Reverse transcriptase			Ribonuclease H		
nucleotide	2309-2839			2840-1-131			132-1163			1164-1625		
	s	r	t	s	r	t	s	r	t	s	r	t
++ 1985 (n 11)*	3	5	6	3	8	8	6	4	6	7	3	8
+- (n 3)*	1	1	2	0	1	1	2	0	2	3	1	3
-- (n 3)*	2	2	2	0	2	2	2	2	3	1	1	2
++ 1998 (n 13)*	7	6	8	5	10	10	10	10	10	10	6	12
Total no. of samples mutated	13	14	18	8	21	21	20	16	21	21	11	25

s = silent mutations; r = replacement mutations; t = the total number of samples with mutations (both s and r, some samples may be counted twice). *n = total number of samples for each particular clinical group. The total number of samples is 30.

Overall the samples have the most nucleotide changes in the Ribonuclease domain, especially on a per nucleotide basis, but the majority of those changes are silent. The last

310 nucleotides of the Ribonuclease domain overlap with the X gene region. The number of samples with mutations in the terminal protein domain is split just about evenly between silent and replacement mutations (s/r ratio =0.93). There are over twice the number of samples with replacement mutations than silent mutations in the spacer domain (s/r ratio = 0.38). There are more samples with silent mutations than replacement mutations in the Reverse transcriptase domain (s/r ratio = 1.25). There are nearly twice the number of samples with silent mutations than replacement mutations in the Ribonuclease H domain (s/r ratio = 1.91).

4.1.6 1985 and 1998 comparisons

Samples where both 1998 and 1985 genomic sequences were available were aligned in Se-Al, then exported into PAUP*4, where the distance ratio was compared between 1985 and 1998 samples (table 4.13). All the samples compared belong to the ++ clinical type, nine were genotype D, and two were genotype C.

Table 4.13 the number of single base-pair differences and indels observed between 1998 and 1985 transcripts of the same sample.

Sample	number of differences	Number of deletions (no. of nucleotides)	No. of characters included (%)*
AN12	6		2701 (84)
BU12	1		2640 (82)
CH02	4		3176 (99)
HA53	4		2864 (89)
MO59	4		3181 (99)
PA31	8		2839 (88)
PA58 C	2		1396 (43)
SA18	14	2 (9 nt, 65 nt)	2386 (74)
TA40	0		3153 (98)
TE55 C	0	1 (9 nt)	2372 (74)
WA13	15	3 (54 nt, 36 nt, 11 nt)	3038 (94)

C indicates genotype C, all other samples are genotype D. *percent of total sequence.

The mean number of mutations per sample per year is 1.31×10^{-4}. The number of differences observed between the 1985 and 1998 transcripts range from none at all (TA40 and TE55) to 15 differences (WA13). This analyses effectively only takes in to account

the single nucleotide differences. For several of the samples there are deletions that have occurred in the 1998 samples that are not in the 1985 samples. For example, WA13_1998 has three separate deletions (54 nt, 36 nt and 11 nt). They all occur within a 226 base pair region within the preSI and preSII areas, which overlaps with the variable spacer region of the P gene region. SA18_1998 has a deletion (65 nt) within this general area too. Both SA18_1998 and SA18_1985 have a small deletion, only 9 nt, from nt 1767. TE55_1998 also has a 9 nt deletion in this general area (from nt 1760). This is close to where the nick in the plus strand of DNA is situated.

Overall WA13 and SA18 samples show the most divergence from the 1985 to the 1998 samples, as they have the most single nucleotide changes, and both samples have deletions. One deletion in SA18 was shared in both the 1985 and 1998 transcripts. TE55_1998 shows no nucleotide differences from TE55_1985 for the sequence analysed, yet TE55_1998 has a small deletion that did not appear in the TE55_1985 sequence.

4.2 HLA results

The HLA results will be described in two sections, first analysis of the Kawerau samples, then comparing the Kawerau samples to New Zealand Maori and New Zealand European populations.

4.2.1 Kawerau HLA samples

To investigate if there is a correlation between immunological status and clinical type each individual was HLA characterised. HLA-A, HLA-B, HLA-C, HLA-DRB1 and HLA-DQB1 loci were typed by the Transplantation Immunology Group, Oxford Transplant Centre, Churchill Hospital, Oxford, United Kingdom. From the results the data was reduced to allele family level for ease of comparison and allele frequency (AF) and genotype frequency (GF) were calculated. Allele frequency is defined as:

$$AF = \frac{P}{2n}$$

Where P = the number of times a particular allele was observed in the whole population; 2n = total number of alleles (Henry et al., 1994).

Genotype frequency is defined as:

$$GF = 1 - \sqrt{1 - AF}$$

Where AF is the allele frequency (Henry et al., 1994).

- The most common allele family for HLA-A is *02, with a frequency of 46.9%.
- The most common allele family for HLA-B is *55, with a frequency of 28.6%.
- The most common allele family for HLA-C is *01 with 30.6%.
- The most common allele family for DRB1 is *12, with a frequency of 30.6%.
- DQB1 is the least variable of the HLA genes, with only four allele families, of which *03 was the most frequent, with a frequency of 61.7%.

Table 4.14 Allele family frequency for each HLA gene.

Gene	Allele Family	Number of samples	Allele Frequency
HLA-A	*02	46	0.469
(n (total) = 98)	*24	22	0.224
	*11	12	0.122
	*01	7	0.071
	*03	3	0.031
	*34	3	0.031
	*29	3	0.031
	*26	1	0.010
	*68	1	0.010
HLA-B	*55	28	0.286
(n (total) = 98)	*40	15	0.153
	*48	12	0.122
	*44	10	0.102
	*39	9	0.092
	*07	5	0.051
	*08	4	0.041
	*56	4	0.041
	*35	2	0.020
	*53	2	0.020
	*18	2	0.020
	*45	1	0.010
	*27	1	0.010
	*58	1	0.010
	*14	1	0.010
	*15	1	0.010
HLA-C	*01	30	0.306
(n (total) = 98)	*07	18	0.184
	*04	17	0.173
	*08	12	0.122
	*03	5	0.051
	*05	5	0.051
	*15	3	0.031
	*12	3	0.031
	*02	2	0.020
	*16	2	0.020
	*06	1	0.010
DRB1	*12	30	0.306
(n (total) = 98)	*11	21	0.214
	*14	9	0.092
	*15	8	0.082
	*04	7	0.071
	*03	5	0.051
	*07	5	0.051
	*09	4	0.041
	*08	4	0.041
	*16	2	0.020
	*10	1	0.010
	*01	1	0.010
	*13	1	0.010
DQB1	*03	58	0.617
(n (total) = 94)	*05	15	0.160
	*06	11	0.117
	*02	10	0.106

To see if there was any specific pattern of HLA alleles with regard to hepatitis B virus, the allele frequencies for each HLA gene were broken up within the sample clinical groups for HBV infection (table 4.15).

Table 4.15 HLA allele family frequencies for each of the clinical states of HBV infection.

	HLA-A			HLA-B			HLA-C			DRB1			DQB1		
++	allele	no.	AF Frq	allele	no.	AF Frq	allele	no.	AF Frq	allele	no.	AF Frq	allele	no.	AF Frq
	*02	15	0.577	*55	7	0.269	*01	7	0.269	*11	8	0.308	*03	16	0.615
	*24	6	0.231	*40	5	0.192	*04	7	0.269	*12	7	0.269	*05	4	0.154
	*11	2	0.077	*48	4	0.154	*07	4	0.154	*15	3	0.115	*06	3	0.115
	*26	1	0.038	*39	2	0.077	*08	3	0.115	*16	2	0.077	*02	3	0.115
	*03	1	0.038	*08	2	0.077	*02	2	0.077	*14	2	0.077			
	*01	1	0.038	*07	2	0.077	*12	1	0.038	*03	2	0.077			
				*45	1	0.038	*06	1	0.038	*07	1	0.038			
				*44	1	0.038	*03	1	0.038	*04	1	0.038			
				*35	1	0.038									
				*27	1	0.038									
+-	*02	13	0.464	*55	10	0.385	*01	11	0.393	*12	11	0.393	*03	20	0.714
	*24	7	0.25	*40	4	0.143	*07	6	0.214	*11	6	0.214	*05	4	0.143
	*34	2	0.071	*44	3	0.107	*04	3	0.107	*14	3	0.107	*02	2	0.071
	*29	2	0.071	*48	3	0.107	*08	3	0.107	*04	3	0.107	*06	2	0.071
	*11	2	0.071	*39	3	0.107	*05	2	0.071	*07	2	0.071			
	*03	1	0.036	*58	1	0.036	*16	1	0.036	*15	1	0.036			
	*01	1	0.036	*56	1	0.036	*15	1	0.036	*09	1	0.036			
				*18	1	0.036	*03	1	0.036	*08	1	0.036			
				*16	1	0.036									
				*07	1	0.036									
--	*02	18	0.429	*55	11	0.262	*01	12	0.286	*12	12	0.286	*03	21	0.553
	*24	9	0.214	*44	6	0.143	*07	8	0.190	*11	6	0.143	*05	6	0.158
	*11	7	0.167	*48	5	0.119	*04	7	0.167	*14	4	0.095	*06	6	0.158
	*01	4	0.095	*40	5	0.119	*08	6	0.143	*15	3	0.071	*02	5	0.132
	*68	1	0.024	*39	4	0.095	*05	3	0.071	*09	3	0.071			
	*34	1	0.024	*56	3	0.071	*03	3	0.071	*08	3	0.071			
	*29	1	0.024	*53	2	0.048	*16	1	0.024	*04	3	0.071			
	*03	1	0.024	*08	2	0.048	*15	1	0.024	*03	3	0.071			
				*07	2	0.048	*12	1	0.024	*07	2	0.048			
				*35	1	0.024				*13	1	0.024			
				*14	1	0.024				*10	1	0.024			
										*01	1	0.024			

- HLA-A*02 and HLA-A*24 were consistently the most frequently occurring alleles for HLA-A. In the -- clinical type, there is a higher frequency of *11 and *01 alleles compared to the other clinical types.
- For the HLA-B gene, the *55 allele family is the most frequent for all clinical groups. The *44 allele had a slightly lower proportion in the ++ clinical group compared to the others.
- The most frequent alleles from the HLA-C gene are *01, *07, *04, and *08, this pattern is consistent for all clinical types.

- Alleles DRB1*11 and *12 are consistently the most frequent alleles for the DRB1 gene.
- For the DQB1 gene, *03 is the most frequent allele, occurring over 55% in each clinical type.

The frequency of specific genotypes for each HLA gene were compared between each (Table 4.16).

Table 4.16 Genotype frequencies for each of the clinical states.

	HLA-A			HLA-B			HLA-C			DRB1			DQB1		
++	Genotype	no.	Freq	Genotype	no.	Freq	Genotype	no.	Freq	Genotype	no.	Freq	Genotype	no.	Freq
	*02*24	5	0.385	*40*48	2	0.153	*04*04	2	0.153	*11*12	3	0.231	*03*03	7	0.538
	*02*02	3	0.231	*40*55	2	0.153	*01*01	2	0.153	*11*11	2	0.153	*02*06	1	0.077
	*02*26	1	0.077	*45*40	1	0.077	*07*07	1	0.077	*15*07	1	0.077	*02*02	1	0.077
	*02*03	1	0.077	*08*08	1	0.077	*07*12	1	0.077	*03*03	1	0.077	*05*06	1	0.077
	*02*11	1	0.077	*07*39	1	0.077	*02*08	1	0.077	*15*16	1	0.077	*06*03	1	0.077
	*11*24	1	0.077	*27*48	1	0.077	*01*07	1	0.077	*04*11	1	0.077	*05*03	1	0.077
	*02*01	1	0.077	*07*55	1	0.077	*01*04	1	0.077	*15*12	1	0.077	*05*05	1	0.077
				*48*55	1	0.077	*01*08	1	0.077	*12*14	1	0.077			
				*55*39	1	0.077	*03*06	1	0.077	*16*14	1	0.077			
				*35*44	1	0.077	*04*08	1	0.077	*12*12	1	0.077			
				*55*55	1	0.077	*02*04	1	0.077						
+-	*02*24	3	0.214	*40*55	2	0.143	*01*07	3	0.214	*12*12	3	0.214	*03*03	8	0.571
	*02*34	2	0.143	*55*55	2	0.143	*01*04	2	0.143	*04*12	2	0.143	*05*03	3	0.214
	*02*29	2	0.143	*39*56	1	0.071	*01*01	2	0.143	*12*14	1	0.071	*02*06	1	0.071
	*02*02	2	0.143	*07*18	1	0.071	*07*07	1	0.071	*04*11	1	0.071	*02*05	1	0.071
	*03*24	1	0.071	*44*55	1	0.071	*01*05	1	0.071	*14*12	1	0.071	*03*06	1	0.071
	*24*11	1	0.071	*39*55	1	0.071	*08*08	1	0.071	*11*11	1	0.071			
	*01*02	1	0.071	*48*48	1	0.071	*05*16	1	0.071	*15*07	1	0.071			
	*02*11	1	0.071	*44*44	1	0.071	*03*15	1	0.071	*14*07	1	0.071			
	*24*24	1	0.071	*58*55	1	0.071	*01*08	1	0.071	*09*11	1	0.071			
				*40*15	1	0.071	*04*07	1	0.071	*11*12	1	0.071			
				*39*40	1	0.071				*09*11	1	0.071			
				*48*55	1	0.071									
--	*02*24	4	0.190	*39*55	3	0.143	*01*07	4	0.190	*12*14	2	0.095	*03*03	6	0.316
	*02*02	4	0.190	*44*55	3	0.143	*03*07	2	0.095	*04*12	2	0.095	*05*03	5	0.263
	*02*11	4	0.190	*44*48	2	0.095	*04*04	2	0.095	*12*12	2	0.095	*02*06	3	0.158
	*01*01	1	0.048	*40*55	2	0.095	*01*01	2	0.095	*04*11	1	0.048	*06*03	2	0.105
	*24*11	1	0.048	*08*40	1	0.048	*01*05	1	0.048	*03*09	1	0.048	*02*03	2	0.105
	*03*24	1	0.048	*39*56	1	0.048	*05*08	1	0.048	*07*11	1	0.048	*05*06	1	0.053
	*02*68	1	0.048	*35*40	1	0.048	*01*12	1	0.048	*10*11	1	0.048			
	*24*34	1	0.048	*55*56	1	0.048	*04*08	1	0.048	*01*12	1	0.048			
	*24*24	1	0.048	*48*40	1	0.048	*08*15	1	0.048	*08*14	1	0.048			
	*11*29	1	0.048	*07*07	1	0.048	*04*01	1	0.048	*14*09	1	0.048			
	*01*02	1	0.048	*08*55	1	0.048	*03*16	1	0.048	*08*12	1	0.048			
	*01*11	1	0.048	*53*53	1	0.048	*04*07	1	0.048	*15*07	1	0.048			
				*48*55	1	0.048	*01*08	1	0.048	*15*03	1	0.048			
				*44*56	1	0.048	*05*07	1	0.048	*03*15	1	0.048			
				*14*48	1	0.048	*08*08	1	0.048	*09*12	1	0.048			
										*11*08	1	0.048			
										*11*12	1	0.048			
										*11*13	1	0.048			

For all the HLA genes and all clinical types, finding a genotype that is the more frequent than others was not a foregone conclusion, except in the DQB1 gene, where the *03*03 genotype occurs more frequently than the other genotypes.

The haplotype frequencies of each HLA gene (see section 1.2.3 in Chapter 1 for explanation on haplotypes) were also compared (see appendix B for haplotype tables of Kawerau data). Chi square tests were performed on the haplotype data, then checked with Monte Carlo Estimates of the exact Chi square test using SAS software. The total number of samples was low, which makes the Chi square test less accurate, that is why the Monte Carlo estimates were performed. The results are summarised Table 4.17.

Table 4.17 Chi square results for the haplotype data.

HLA haplotype	Clinical type	*p*	Monte Carlo estimate	Significance
AB	all	0.0004	0.0223	*
AB	++	0.003	0.0196	*
AB	+-	0.001	0.0061	**
AB	--	0.0276	0.0647	-
AC	all	<0.0001	<0.0001	**
AC	++	0.0408	0.1002	-
AC	+-	0.2069	0.2576	-
AC	--	<0.0001	0.0041	**
CB	all	<0.0001	<0.0001	**
CB	++	0.0006	0.0030	**
CB	+-	<0.0001	<0.0001	**
CB	--	0.0013	0.0041	**
DQB1 DRB1	all	<0.0001	<0.0001	**
DQB1 DRB1	++	<0.0001	<0.0001	**
DQB1 DRB1	+-	0.0005	0.0066	**
DQB1 DRB1	--	<0.0001	<0.0001	**

NB * is significant; ** is highly significant, - is not significant. AB = HLA-A and HLA-B; AC = HLA-A and HLA-C; CB = HLA-B and HLA-C; DQB1 DRB1 = HLA-DQB1 and HLA-DRB1.

For all clinical types together, the HLA allele haplotypes show linkage disequilibrium which is expected in populations with minimal diversity. The dominant alleles from each HLA gene are inherited together as haplotypes.

4.2.2 Comparing Kawerau samples with other New Zealand Populations.

The Kawerau samples are of either Maori or Polynesian origin. In order to compare the data for the general New Zealand populations, which is only available in serological form (Dagger et al., 1986; Fong et al., 1981; Henry et al., 1994), the data from this study was converted from sequence based typing (SBT) to the serological names. Table 4.18 serves as a conversion table between serological names and SBT names (Bodmer et al., 1999). HLA allele frequencies for all HLA genes analysed were compared to data from New Zealand Maori and New Zealand European populations, both these sample sets were from the general population (tables 4.19 to 4.23), in each table the most frequent alleles are shaded.

Table 4.18 Table converting the serological names for HLA alleles to SBT equivalents.

HLA gene	Serological name	SBT equivalent(s)
HLA-A	A9	A*24
	A10	A*26, A*34
	A19	A*29, A*30, A*31, A*32, A*33
	A28	A*68
HLA-B	B5	B*51, B*52
	B12	B*44, B*45
	B16	B*39
	B17	B*57, B*58
	B21	B*49, B*50
	B22	B*55, B*56
HLA-DRB1	DR2	DRB1*15, DRB1*16
	DR5	DRB1*11, DRB1*12
HLA-DQB1	DQw1	DQB1*5, DQB1*6

NB only the names that are new to sequence based typing (SBT) and that are relevant to these data sets have been included.

Table 4.19 HLA-A allele frequencies for New Zealand populations.

Allele	NZ Maori		NZ European		Kawerau	
HLA-A	n (244)	Freq	n (3494)	Freq	n (98)	Freq
A1	16.1	0.066	590	0.16	7	0.071
Aw36	0.5	0.002	0	0	0	0
A2	75	0.307	847	0.242	46	0.469
A28	4	0.016	96	0.027	1	0.01
A3	12.3	0.05	426	0.122	3	0.031
A11	39.1	0.16	239	0.068	12	0.128
A9	63.6	0.261	335	0.096	22	0.224
A10	16.1	0.066	161	0.046	4	0.041
A19	9.2	0.038	344	0.098	3	0.031
Blank	8.2	0.034	461	0.132	0	0

All three groups (NZ Maori, NZ European, and Kawerau) have HLA-A2 as their dominant allele. Both Maori groups (NZ Maori and Kawerau) have A9 and A11 as their next most frequent alleles, whereas NZ Europeans have A3 and A1 as their next most common alleles. A3 and A1 both occur at a frequency equal or less than 0.08 for both Maori groups. The differences in allele frequencies seen between the Maori groups and the NZ European group were expected. Any differences in allele frequencies between the Maori and Kawerau groups may be due to the Kawerau group all having HBV infection.

Table 4.20 HLA-B allele frequencies for New Zealand populations.

Allele	NZ Maori		NZ European		Kawerau	
HLA-B	n (244)	Freq	n (3494)	Freq	n (98)	Freq
B5	7.1	0.029	157	0.045	0	0
B35	6.5	0.027	231	0.066	2	0.02
B15	7.1	0.029	211	0.06	1	0.01
B17	3.5	0.014	169	0.048	1	0.01
B27	4.5	0.018	142	0.041	1	0.01
B16	24.8	0.101	93	0.027	9	0.092
B22	41.8	0.172	114	0.033	32	0.327
B13	4	0.016	70	0.02	0	0
B40	52.6	0.216	276	0.079	15	0.153
B48	16.4	0.067	2	0.0006	12	0.122
B8	13.7	0.056	430	0.123	4	0.041
B14	5.1	0.021	122	0.035	1	0.01
B37	1.5	0.006	44	0.013	0	0
B12	17.2	0.07	583	0.167	11	0.112
B18	0	0	100	0.029	2	0.02

B21	0	0	52	0.015	0	0
B41	0	0	9	0.003	0	0
B42	0	0	2	0.0006	0	0
B47	0	0	3	0.0009	0	0
B53	0	0	0	0	2	0.02
Blank	19.5	0.08	230	0.066	0	0

B22 is the most frequent allele in the Kawerau samples, occurring at a rate twice as much as the next most common allele (B40). The most common allele for the NZ Maori group is B40, with B22 second. Both Maori groups have high frequencies of B48 (0.067, and 0.122) compared to the NZ European group, which has B48 at a frequency of 0.0006.

Comparing the two Maori groups, they are again reasonably similar, except the Kawerau group have a frequency of 0.112 for B12, which is almost double the frequency for B12 in the general Maori population (0.07). B12 is split into B*44 and B*45 in the sequence based typing naming, counting B*44 and B*45 together for the Kawerau data, there are more of these alleles in the −− clinical group that the other clinical groups (6−−; 3+−; 2++). Possibly the presence of this HLA-B allele benefits the hosts with this allele, as over half the individuals with this allele have low viral titre.

Table 4.21 HLA-C allele frequencies for New Zealand populations.

Allele	NZ Maori		NZ European		Kawerau	
HLA-C	n (144)	Freq	n	Freq	n (98)	Freq
*01	27	0.188	Data		30	0.306
*02	2.6	0.018	not		2	0.02
*03	3.5	0.024	available		5	0.051
*04/*06	14.7	0.102			18	0.184
*05	8.3	0.058			5	0.051
*07	15.9	0.11			18	0.184
*08	0	0			12	0.122
*15	0	0			3	0.031
*12	0	0			3	0.031
*16	0	0			2	0.02
Blank	72	0.5			0	0

There was no NZ European data available for the remaining HLA genes. Both Maori groups again show similar allele frequencies, except the Kawerau group has a high frequency of C*08, this allele does not occur in the general Maori population. Several

other alleles (C*15, C*12, C*16) occur in the Kawerau group but do not occur in the general Maori population. Possibly the Kawerau samples have more admixture with other ethnic groups than the general Maori population.

Table 4.22 HLA-DRB1 allele frequencies for New Zealand populations.

Allele	NZ Maori		NZ European		Kawerau	
HLA-DRB1	n (98)	Freq	n	Freq	n (98)	Freq
DR1	4.1	0.042	Data		1	0.01
DR10	0	0	not		1	0.01
DR2	7.3	0.074	available		10	0.102
DR3	4.6	0.047			5	0.051
DR6	10.1	0.103			10	0.102
DR5	31.1	0.317			51	0.52
DR8	8.9	0.091			4	0.041
DR4	17.7	0.181			7	0.071
DR7	4.6	0.047			5	0.051
DR9	7.3	0.074			4	0.041
Blank	2.3	0.023			0	0

The most common allele in both the NZ Maori and Kawerau groups is DR5. DR5 accounts for over 50% of the alleles in the Kawerau group, and accounts for just over 30% in the general Maori population. Both of these sample sizes are small, so the slight variations seen in allele frequencies may be due to the small sample size.

Table 4.23 HLA-DQB1 allele frequencies for New Zealand populations.

Allele	NZ Maori		NZ European		Kawerau	
HLA-DQB1	n (98)	Freq	n	Freq	n (94)	Freq
DQw1	29.6	0.302	Data		26	0.277
DQw2	6.8	0.069	not		10	0.106
DQw3	57.2	0.584	available		58	0.617
Blank	4.4	0.045			0	0

In general all the Kawerau DQB1 allele frequencies correlate very well with the general NZ Maori population, with DQw3 being the most frequent allele with a frequency of about 60% for both groups. The haplotype frequencies of each HLA gene combination samples were compared to other New Zealand populations who were not tested specifically for hepatitis B (table 4.24).

Table 4.24 Haplotype frequency for NZ Maori and NZ European and Kawerau samples

Maori Samples n=576		European Samples n=1747		Kawerau Samples n=46	
HLA-AB	hap. freq	HLA-AB	hap. freq	HLA-AB	hap. freq
A2 B22	14.8	A2 B22	0.8	A2 B22	19.6
A9 B22	4.7	A9 B22	0.7	A9 B22	6.5
A2 B12	2.2	A2 B12	8.1	A2 B12	6.0
A2 B48	<0.2	A2 B48	-	A2 B48	5.4
A2 B40	2.5	A2 B40	2.8	A2 B40	4.9
A2 B16	1.0	A2 B16	0.5	A2 B16	4.3
A9 B40	12.3	A9 B40	1.5	A9 B40	3.3
A9 B48	0.5	A9 B48	-	A9 B48	3.3
A9 B16	6.7	A9 B16	0.9	A9 B16	2.7
A2 B7	1.8	A2 B7	4.4	A2 B7	2.7
A11 B22	1.6	A11 B22	1.0	A11 B22	2.7
A11 B48	0.5	A11 B48	<0.2	A11 B48	2.7
A11 B40	12.2	A11 B40	0.8	A11 B40	2.2
A10 B40	3.5	A10 B40	0.4	A10 B40	2.2
A19 B12	1.5	A19 B12	3.6	A19 B12	2.2
A1 B8	4.3	A1 B8	10.3	A1 B8	1.6
A11 B12	0.3	A11 B12	1.2	A11 B12	1.6
A3 B7	2.3	A3 B7	5.4	A3 B7	1.1
A9 B8	0.3	A9 B8	0.4	A9 B8	1.1
A2 B53	-	A2 B53	-	A2 B53	1.1
A9 B12	0.7	A9 B12	1.6	A9 B12	1.1
A1 B22	0.3	A1 B22	0.4	A1 B22	1.1
A1 B40	0.6	A1 B40	0.4	A1 B40	1.1
A1 B7	<0.2	A1 B7	1.3	A1 B7	1.1
A19 B22	-	A19 B22	<0.2	A19 B22	1.1
A9 B53	-	A9 B53	-	A9 B53	1.1
A2 B35	0.6	A2 B35	1.1	A2 B35	0.5
A2 B27	0.5	A2 B27	1.5	A2 B27	0.5
A2 B17	0.6	A2 B17	1.5	A2 B17	0.5
A2 B14	0.4	A2 B14	0.7	A2 B14	0.5
A2 B15	1.1	A2 B15	3.0	A2 B15	0.5
A9 B7	0.7	A9 B7	1.4	A9 B7	0.5
A9 B35	0.3	A9 B35	0.9	A9 B35	0.5
A9 B18	<0.2	A9 B18	<0.2	A9 B18	0.5
A11 B8	<0.2	A11 B8	0.3	A11 B8	0.5
A11 B27	0.3	A11 B27	0.3	A11 B27	0.5
A11 B14	<0.2	A11 B14	<0.2	A11 B14	0.5
A1 B12	<0.2	A1 B12	1.3	A1 B12	0.5
A1 B35	0.7	A1 B35	0.8	A1 B35	0.5
A1 B17	1.1	A1 B17	3.0	A1 B17	0.5
A3 B40	0.3	A3 B40	0.9	A3 B40	0.5
A3 B16	-	A3 B16	<0.2	A3 B16	0.5
A3 B35	0.4	A3 B35	1.6	A3 B35	0.5
A3 B18	<0.2	A3 B18	0.7	A3 B18	0.5
A10 B22	1.0	A10 B22	<0.2	A10 B22	0.5
A10 B48	-	A10 B48	-	A10 B48	0.5
A10 B15	<0.2	A10 B15	<0.2	A10 B15	0.5
A10 B12	<0.2	A10 B12	0.8	A10 B12	0.5
A28 B22	<0.2	A28 B22	<0.2	A28 B22	0.5

A1 B5	<0.2	A1 B5	0.4	A1 B5	-
A1 B15	<0.2	A1 B15	0.8	A1 B15	-
A1 B37	0.3	A1 B37	0.6	A1 B37	-
A2 B5	1.1	A2 B5	1.6	A2 B5	-
A2 B8	0.6	A2 B8	1.5	A2 B8	-
A2 B13	0.3	A2 B13	0.5	A2 B13	-
A2 B18	0.3	A2 B18	0.6	A2 B18	-
A2 B21	<0.2	A2 B21	0.8	A2 B21	-
A2 B37	-	A2 B37	0.3	A2 B37	-
A3 B5	<0.2	A3 B5	0.5	A3 B5	-
A3 B8	<0.2	A3 B8	0.4	A3 B8	-
A3 B14	0.3	A3 B14	1.2	A3 B14	-
A3 B15	0.3	A3 B15	0.9	A3 B15	-
A3 B17	0.3	A3 B17	0.3	A3 B17	-
A3 B27	<0.2	A3 B27	0.5	A3 B27	-
A9 B5	<0.2	A9 B7	0.5	A9 B7	-
A9 B14	<0.2	A9 B14	0.3	A9 B14	-
A9 B15	1.3	A9 B15	1.1	A9 B15	-
A9 B17	<0.2	A9 B17	0.4	A9 B17	-
A9 B21	<0.2	A9 B21	0.4	A9 B21	-
A9 B27	0.3	A9 B27	0.3	A9 B27	-
A10 B5	0.3	A10 B5	<0.2	A10 B5	-
A10 B7	<0.2	A10 B7	0.3	A10 B7	-
A10 B14	-	A10 B14	0.3	A10 B14	-
A10 B16	0.5	A10 B16	0.6	A10 B16	-
A10 B17	0.3	A10 B17	<0.2	A10 B17	-
A10 B18	<0.2	A10 B18	0.8	A10 B18	-
A10 B27	<0.2	A10 B27	0.3	A10 B27	-
A10 B35	<0.2	A10 B35	0.3	A10 B35	-
A11 B5	<0.2	A11 B5	0.6	A11 B5	-
A11 B7	1.1	A11 B7	0.7	A11 B7	-
A11 B15	<0.2	A11 B15	0.3	A11 B15	-
A11 B16	0.9	A11 B16	<0.2	A11 B16	-
A11 B35	0.4	A11 B35	1.6	A11 B35	-
A19 B5	<0.2	A19 B5	0.6	A19 B5	-
A19 B7	0.3	A19 B7	0.4	A19 B7	-
A19 B8	<0.2	A19 B8	0.4	A19 B8	-
A19 B13	<0.2	A19 B13	0.8	A19 B13	-
A19 B14	-	A19 B14	0.7	A19 B14	-
A19 B15	-	A19 B15	0.4	A19 B15	-
A19 B17	-	A19 B17	0.3	A19 B17	-
A19 B18	<0.2	A19 B18	0.5	A19 B18	-
A19 B27	0.3	A19 B27	0.8	A19 B27	-
A19 B35	0.3	A19 B35	0.4	A19 B35	-
A19 B40	0.5	A19 B40	1.2	A19 B40	-
A28 B12	0.4	A28 B12	0.9	A28 B12	-
A28 B40	0.3	A28 B40	0.4	A28 B40	-

NB Only allele haplotype frequencies above 0.2 are shown for the non- Kawerau samples. - indicates haplotype does not occur. The three most frequent haplotypes are shaded.

Both the NZ Maori and NZ European populations have more haplotypes than the Kawerau samples, this is indicative of the much larger sample sizes (576 and 1747, respectively,

compared to 46). Both the NZ Maori and Kawerau samples have A2 B22 as their most frequent haplotype, in comparison the highest haplotype frequency for the NZ European samples is A1 B8. Haplotype A2 B22 occurs at a frequency of less than 1.0 in NZ Europeans.

The Kawerau samples have a high frequency of the A2 B48 haplotype (frequency = 5.4), whereas the NZ Maori population in general only has this haplotype at a frequency of less than 0.2. This haplotype is not in the NZ European population at all. The fact that the general Maori population has this haplotype at such a low frequency makes this haplotype a good choice of haplotype that may correlate with the presence of HBV. More populations with HBV will have to be HLA typed to see if this haplotype appears in other populations with HBV. The haplotypes A9 B48 and A11 B48 also show the pattern of being reasonably frequent in the Kawerau samples, but low in the general NZ Maori population; and low, or absent in NZ European populations. The presence of the allele B48 in all these haplotypes suggests that B48 may be important for the presence of HBV. The B48 allele is split reasonably evenly between the clinical groups of the Kawerau samples (++ 4; +- 3; -- 5).

Chapter 5 Summary and Conclusions

5.1 Summary of HBV results

All samples were found to be either Genotype-D or Genotype-C. The genomes sequenced come from viruses that exist in a quasispecies state, therefore there are many other variants in the host. The quasispecies state was highlighted by multiple bands identified when the gene regions were PCR amplified (see Figure 4.1). Thus some virions have insertions or deletions in their genomes resulting in the different banding patterns observed. The quasispecies distribution allows mutations to survive because there are other virions in the host that provide the necessary viral machinery. The mutations themselves are immune escape mutations, and are often selected for because they escape discovery by the host immune system. For all gene regions, more 1998 samples were mutated compared to 1985 samples. Thus the virus incurred mutations over time.

Mutational 'hotspots' are regions that accumulate mutations more than other regions. The most obvious mutational hotspot for these samples was the spacer region of the P gene, which overlaps with preSI and preSII regions of the surface gene (nucleotides 2840-1-137). This region not only had many single nucleotide changes and degenerative sites, but in some samples, deletions. A lack of regions of functional significance means that these deletions are tolerated.

Two other samples had deletions within the X gene region. Both deletions occurred in an area that has three functionally important overlapping areas; these were the functional domain for translation, enhancer II, and the core promoter. The first deletion was 8-nt in both the SA18 samples (nt 1765 to 1773). The other deletion was 9-nt and was situated fractionally upstream (nt 1760 to 1769) and occurred in TE55_1998 only. There is a reported 8-nt deletion that occurs in this area (nt1770 to 1777) which truncates 20 amino acids from the carboxyl terminus of the X protein (Uchida et al., 1997). This deletion has been reported to lead to the suppression of replication and expression of HBV DNA (Uchida et al., 1997), yet both samples with this deletion are from the ++ clinical group –

which is positive for HBeAg and is often a marker for HBV DNA expression. Possibly the deletions that are reported here do not have the same effect as the deletion reported in Uchida *et. al.* (1997). Or other members of the quasispecies do not have this deletion, and thus HBeAg is continued to be expressed.

Four samples (the -- HI43_1985, and the ++ HA53_1998, ++ RA37_1998 and ++ SA18_1998) had the double mutation at nt 1762 and 1764, which results in two amino acid changes, and the loss of nuclear binding receptor site that suppresses the transcription of precore and core mRNA. This double mutation is effectively an immune escape mutation, as the reduction of transcription of precore and core mRNA in turn reduces the expression of the precore and core proteins. Overall not many samples showed mutations in the X-gene region, this is indicative of the many regions of importance to the virus that are found here. Thus this region cannot endure many mutations.

The precore/core gene regions also have many areas of significance. The first I shall discuss is the well documented 1896 immune escape/encapsidation signal mutation. Only two samples had the immune escape mutation that occurs at nt 1896. This mutation prevents the expression of the HBeAg. Yet, one of those samples belonged to the ++ clinical group (RA37_1998). This anomaly is due to the quasispecies distribution. The virion that was sequenced from this sample had a mutation which prevents the expression of the HBeAg, yet serological tests for this sample came back as HBeAg positive. Clearly there are other virions in this host that do not have this mutation, and are producing the HBeAg. The other sample with this mutation belonged to the -- clinical group.

The epitopes of the virus are areas which interact with the host immune system, thus from the virus perspective, it is a "good idea" to stop these interactions. The precore/core and surface genes are the areas with epitope regions. Many (11 out of 19) of the samples had mutations within these epitope regions.

By comparing the 1985 and 1998 data, I calculated the mean number of mutations per sample per year to be 1.31×10^{-4} within a single host. Another mutation rate has been

estimated that at a single nucleotide position of HBV $1.3x10^{-5}$ mutations occur in an infected individual each year (Carmen and Thomas, 1992; Okamoto et al., 1987a). The difference between these mutation-rate estimates could be due to differences of HBV, such as different genotypes or different clinical types of HBV infection. Although HBV seems to be highly evolved, it still has the capacity for further diversity. A number of the reported variations, both in this study and in others, have an obvious clinical relevance, but others do not, and further study is required to elucidate their importance. In some studies it remains to be determined whether these variations represent independently transmissible strains of HBV or are all selected during the course of infection (that is within a single host), only to be lost in favour of the original strain on infection of another host. I expect that the mutation rate of HBV is higher within a single host, compared to the mutation rate of HBV that has been transferred between hosts. Many of the mutations that I found will not be passed on to another host. Not all members of a quasi-species are equal in this respect.

5.2 HLA results summary

The HLA typing on the human hosts showed a distribution of frequencies expected for a Polynesian population. For example for HLA-A gene, the HLA-A*2 allele is the most common. The most common alleles were similar for each clinical group. Differences between clinical groups were few, but one allele in particular did stand out. This was HLA-B12 (actually HLA-B*44 and HLA-B*45 but they became B12 by serological naming). This allele occurred at relatively low frequencies in the ++ clinical type, compared to the other clinical types. But when comparing the Kawerau group to the NZ Maori population in general, the Kawerau group had nearly double the frequency for this allele than the general Maori population. When haplotype frequencies were compared, any haplotype that contained this allele occurred at a high frequency in the Kawerau samples, but was found at low frequencies for both the NZ Maori and NZ European populations. This suggests that this allele may be correlated with the presence of HBV infection.

5.3 Conclusions

Overall, this thesis provides a very promising start on following hepatitis B infection over time. The 1998 samples had acquired more mutations compared to the 1985 samples, supporting the hypothesis that HBV incurs mutations over time. A comparison of 1985/1998 samples shows that the viruses that remain active (++ and +− clinical groups) gain more mutations than the viruses that become dormant (−−), but due to low number of samples that were sequenced for the +− and −− clinical groups, this result is not conclusive. There are some HLA alleles that correlate with HBV infection. But further work needs to be carried out to confirm this finding.

Bibliography

Arauz-Ruiz, P., Norder, H., Visona, K., and Magnius, L. (1997). Genotype F Prevails in HBV Infected Patients of Hispanic Origin in Central America and May Carry the Precore Stop Mutant., Journal of Medical Virology *51*, 305-312.

Bancroft, W., Mundon, F., and Russell, P. (1972). Detection of additional antigenic determinants of hepatitis B antigen., Journal of Immunology *109*, 842-848.

Bartenschlarger (1990). The P Gene Product of Hepatitis B Virus is Required as a Structural Component for Genomic RNA Encapsidation., Journal of Virology *64*, 5324-5332.

Becker, S., Lee, T.-H., Butel, J., and Slagle, B. (1998). Hepatitis B Virus X Protein Interferes with Cellular DNA Repair., Journal of Virology *72*, 266-272.

Benn, J., and Schneider, R. (1994). Hepatitis B virus HBx protein activates Ras-GTP complex formation and establishes a Ras, Raf, MAP kinase signalling cascade., Proceedings of the National Academy of Science *91*, 10350-10354.

Bertoletti, A., Southwood, S., Chestnut, R., Sette, A., Falco, M., Ferrara, G., Penna, A., Boni, C., Fiaccadori, F., and Ferrari, C. (1997). Molecular Features of the Hepatitis B Virus Nucleocapsid T-Cell Epitope 18-27: Interact With HLA and T-Cell Receptor., Hepatology *26*, 1027-1034.

Bodmer, J., Marsh, S., Albert, E., Bodmer, W., Bontrop, R., Dupont, B., Erlich, H., Hansen, J., Mach, B., Mayr, W., *et al.* (1999). Nomenclature for Factors of the HLA System, Vox Sanguinis *77*, 164-191.

Bodmer, W. (1995). Evolution and Function of the HLA Region., Cancer Surveys *22*, 5-16.

Bowyer, S., and Sim, J. (2000). Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates., Journal of General Virology *81*, 379-392.

Bozkaya, H., Akarca, U., Ayola, B., and Lok, A. (1997). High degree of conservation in the hepatitis B virus core gene during the immune tolerant phase in perinatally acquired chronic hepatitis B virus infection., Journal of Hepatology *26*, 508-516.

Bozkaya, H., Ayola, B., and Lok, A. (1996). High Rates of Mutations in the Hepatitis B Core Gene During the Immune Clearance Phase of Chronic Hepatitis B Virus Infection., Hepatology *24*, 32-37.

Brown, J., Carmen, W., and Thomas, H. (1992). The Clinical Significance of Molecular Variation within the Hepatitis B Virus Genome., Hepatology *15*, 144-148.

Bruss, V., Gerhardt, E., Vieluf, K., and Wunderlich, G. (1996). Functions of the Large Hepatitis B Virus Surface Protein in Viral Particle Morphogenesis., Intervirology *39*, 23-31.

Caillat-Zucman, S., Gimenez, J., Wambergue, F., Albouze, G., Lebkiri, B., Naret, C., Moynot, A., Jungers, P., and Bach, J. (1998). Distinct HLA class II alleles determine antibody response to vaccination with hepatitis B surface antigen., Kidney International *53*, 1626-1630.

Carmen, W., Boner, W., Fattovich, G., Colman, K., Dornan, E., Thursz, M., and Hadziyannis, S. (1997). Hepatitis B Virus Core Protein Mutations are Concentrated in B Cell Epitopes in Progressive Disease and in Helper T Cell Epitopes During Clinical Remission., The Journal of Infectious Diseases *175*, 1093-1100.

Carmen, W., and Thomas, H. (1992). Genetic Variation in Hepatitis B Virus., Gastroenterology *102*, 711-719.

Chen, S., La Porte, P., and Yee, J. (1993). Mutational Analysis of Hepatitis B Virus Enhancer 2., Virology *196*, 652-659.

Courouce-Pauty, A., Lemaire, J., and Roux, J. (1978). New hepatitis B surface antigen subtypes inside the *ad* category., Vox Sanguinis *35*, 304-308.

Cross, J., Wen, C., and Rutter, W. (1993). Transactivation by hepatitis B virus X protein is promiscuous and dependent on mitogen-activated cellular serine/threonine kinases., Proceedings of the National Academy of Science *90*, 8078-8082.

Dagger, J., Faed, J., Fong, R., Milligan, L., Roberts, M., Thomas, H., Webster, A., Woodfield, D., and Whyte, J. (1986). HLA Antigens in New Zealand Maori. In Aizawa Med HLA In Asia-Oceania (Sapporo, Hokkaido University Press), pp. 302-307.

De-Medina, T., Haviv, I., Noiman, S., and Shaul, Y. (1994). The X protein of hepatitis B virus has a ribo-deoxy ATPase activity., Virology *202*, 401-407.

del Guercio, M., Sidney, J., Hermanson, G., Perez, C., Grey, H., Kubo, R., and Sette, A. (1995). Binding of a Peptide Antigen to Multiple HLA Alleles Allows Definition of an A2-Like Supertype, The journal of Immunology *154*, 685-693.

Fong, R., Dagger, J., and Locke, H. (1981). Analysis of HLA - ABC and DR Antigens in New Zealand Maori. Paper presented at: The 2nd Asia and Oceania Histocompatibility Workshop Conference (Toorak, Australia, Immunopublishing).

Franco, A., Barnaba, V., Natali, P., Balsano, C., Musca, A., and Balsano, F. (1988). Expression of class I and class II major histocompatibility complex antigens on human hepatocytes., Hepatology *8*, 449-454.

Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P., and Charnay, P. (1979). Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in *E. coli*., Nature *281*, 646-650.

Grob, P. (1998). Hepatitis B: virus, pathogenesis and treatment., Vaccine *16*, S11-S16.

Gunther, S., Fischer, L., Pult, I., Sterneck, M., and Will, H. (1999). Naturally Occurring Variants of Hepatitis B Virus, Advances in Virus Research *52*, 25-137.

Hannoun, C., Horal, P., and Lindh, M. (2000). Lon-term mutations rates in the hepatitis B virus genome., Journal of General Virology *81*, 75-83.

Harrison, G. (1999) Hepatitis B Virus Infection in the South Pacific, MSc, Massey University, Palmerston North.

Henry, S., Perry, H., Roberts, M., and Woodfield, G. (1994). HLA class I gene, antigen and haplotype frequencies in New Zealand Maori and Europeans., The New Zealand Medical Journal *107*, 119-121.

Hohler, T., Gerken, G., Notghi, A., Lubjuhn, R., Taheri, H., Protzer, U., Lohr, H., Schneider, P., Meyer zum Buschenfelde, K., and Rittner, C. (1997). HLA-DRB1*1301 and *1302 protect against chronic hepatitis B., Journal of Hepatology *26*, 503-507.

Horikita, M., Itoh, S., Yamamoto, K., Shibayama, T., Tsuda, F., and Okamoto, H. (1994). Differences in the Entire Nucleotide Sequence Between Hepatitis B Virus Genomes From Carriers Positive for Antibody to Hepatitis B e Antigen With and Without Active Disease., Journal of Medical Virology *44*, 96-103.

Huang, J., Kwong, J., Sun, C.-Y., and Liang, T. (1996). Proteasome complex as a potential cellular target of hepatitis B virus X protein., Journal of Virology *70*, 5582-5591.

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

Lai, M., Solinas, A., Mazzoleni, A., Deplano, A., Farci, P., Lisci, V., Porru, A., Tocco, A., and Balestrieri, A. (1994). The role of pre-core hepatitis B virus mutants on the long-term outcome of chronic hepatitis B virus hepatitis. A longitudinal study., Journal of Hepatology *20*, 773-781.

Le Bouvier, G. (1971). The Heterogeneity of Australia antigen., Journal of Infectious Disease *123*, 671-675.

Lee, T.-H., Elledge, S., and Butel, J. (1995). Hepatitis B virus X protein interacts with a probable cellular DNA repair protein., Journal of Virology *64*, 1107-1114.

Li, J., Buckwold, V., Hon, M., and Ou, J. (1999). Mechanisms of suppression of hepatitis B virus precore RNA transcription by a frequent double mutation., Journal of Virology *73*, 1239-1244.

Liang, T., Baruch, Y., Ben-Porath, E., Enat, R., Bassen, L., Brown, N., Rimon, N., Blum, H., and JR, W. (1991). Hepatitis B virus infection in patients with idiopathic liver disease., Hepatology *13*, 1044-1051.

Lindh, M., Hannoun, C., and Dhillon, A. (1999). Core Promoter Mutations and Genotypes in Relation to Viral Replication and Liver Damage in East Asian Hepatitis B Virus Carriers., Journal of Infectious Disease *179*, 775-782.

Lo, W., and Ting, L. (1994). Repression of enhancer II activity by a negative regulatory element in the hepatitis B virus genome., Journal of Virology *68*, 1758-1764.

Lok, A., Akarca, U., and Greene, S. (1994). Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal., Proceedings of the National Academy of Science *91*, 4077-4081.

Magnius, L., and Norder, H. (1995). Subtypes, Genotypes and Molecular Epidemiology of the Hepatitis B Virus as Reflected by Sequence Vaariability of the S-Gene., Intervirology *38*, 24-34.

Maguire, H., Hoeffler, J., and Siddiqui, A. (1991). HBV X protein alters the DNA binding specificity of CREB and ATF-2 by protein-protein interactions., Science *252*, 842-844.

Mason, W., and Seeger, C. (1991). Current Topics in Microbiology and Immunology, Hepadnaviruses, Molecular Biology and Pathogenesis. (Philadelphia, Springer-Verlag).

Murray-McIntosh, R., Scrimshaw, B., Hatfield, P., and Penny, D. (1998). Testing migration patterns and estimating founding population size in Polynesia by using human mtDNA sequences., Proceedings of the National Academy of Science *95*, 9047-9052.

Natoli, G., Avantaggiati, M., Chirillo, P., Puri, P., Ianni, A., Balsano, C., and Levrero, M. (1994). Ras- and Raf-dependent activation of c-Jun transcriptional activity by the hepatitis B virus transactivator pX., Oncogene *9*, 2837-2843.

Nei, M. (1987). Molecular Evolutionary Genetics (New York, Columbia University Press).

Okamoto, H., Imai, M., Kametani, M., Nakamura, T., and Mayumi, M. (1987a). Genomic heterogeneity of hepatitis B virus in a 54-year-old woman who contracted the infection

through materno-fetal transmission., Japanese Journal of experimental Medicine *57*, 231-236.

Okamoto, H., Imai, M., Tsuda, F., Tanaka, T., Miyaka-wa, Y., and Mayumi, M. (1987b). Point mutations in the S-gene of hepatitis B virus for a *d/y* or *w/r* subtypic change in two blood donors carrying a surface antigen of compound subtype *adyr* or *adwr*., Journal of Virology *61*, 3030-3034.

Okamoto, H., Yotsumoto, S., Akahane, Y., Yamanaka, T., Miyazaki, Y., Sugai, Y., Tsuda, F., Tanaka, T., Miyaka-wa, Y., and Mayumi, M. (1990). Hepatitis B viruses with precore region defects prevail in persistently infected hosts along with seroconversion to the antibody against e antigen., Journal of Virology *64*, 1298-1303.

Palumbi, S., Martin, A., Romano, S., McMillan, W., Stice, L., and Grabowski, G. (1991). The Simple Fools Guide to PCR, Version 2. (Honolulu, University of Hawaii Zoology Department).

Parham, P., and Ohta, T. (1996). Population Biology of Antigen Presentation by MHC Class I Molecules., Science *272*, 67-74.

Pontisso, P., Ruvoletto, M., Gerlich, W., Heermann, K., Bardini, R., and Alberti, A. (1989). Identification of an attachment site for human liver plasma membranes on hepatitis B virus particles., Virology *173*, 522-530.

Pugh, J., and Bassendine, M. (1990). Molecular Biology of hepadnavirus replication. In Viral Hepatitis. British Medical Bulletin, A series of expert reviews., A. Zuckerman, ed. (London, Churchill Livingstone), pp. 329-354.

Qadri, I., Maguire, H., and Siddiqui, A. (1995). Hepatitis B virus transactivator protein X interacts with the TATA-binding protein., Proceedings of the National Academy of Science *92*, 1003-1007.

Radziwill (1990). Mutational Analysis of the Hepatitis B Virus P Gene Product: Domain Structure and RNase H Activity., Journal of Virology *64*, 613-620.

Rambaut, A. (1995). Se-Al (Sequence Alignment Editor) (Oxford).

Ryu, C., Kim, Y., Hur, H., Kim, H., Oh, J., Young, J., and Hong, H. (2000). Mouse Monoclonal Antibodies to Hepatitis B Virus PreS1 Produced After Immunisation with Recombinant PreS1 Peptide., Hybridoma *19*, 185-189.

Saiki, R., Gelfand, D., Stoffel, S., Schare, S., Higuchi, R., Horn, G., Mullis, K., and Erlich, H. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase., Science *239*, 487-491.

Sangfelt, P., Norder, H., Magnius, L., Alaeus, A., Carlsson, T., and Reichard, O. (1997). Interferon-$\alpha_{2\beta}$ treatment in hepatitis B carriers. Effect on hepatitis B virus DNA levels in children infected with different genotypes., Acta Pædiatrica *86*, 135-137.

Schaller, H., and Fischer, M. (1991). Transcriptional Control of Hepadnavirus Gene Expression. In Current Topics in Microbiology and Immunology, W. Mason, and C. Seeger, eds. (Berlin, Springer-Verlag).

Serjeantson, S. W., Ryan, D. P., and Thompson, A. R. (1982). The Colonization of the Pacific: The Story According to Human Leukocyte Antigens, American Journal of Human Genetics *34*, 904-918.

Seto (1990). Transactivation by the hepatitis B virus X protein depends on AP-2 and other transcription factors., Nature *344*, 72-74.

Sette, A., Oseroff, C., Sidney, J., Alexander, J., Chestnut, R., Kakimi, K., Guidotti, L., and Chisari, F. (2001). Overcoming T Cell Tolerance to the Hepatitis B Virus Surface Antigen in Hepatitis B Virus-Transgenic Mice., The journal of Immunology *166*, 1389-1397.

Sitterlin, D., Lee, T.-H., Prigent, S., Tiollais, P., Butel, J., and Transy, C. (1997). Interaction of the UV-damaged DNA-binding protein with hepatitis B virus X protein is conserved among mammalian hepadnaviruses and restricted to transactivation-proficient X-insertion mutants., Journal of Virology *71*, 6194-6199.

Smith, D., McAllister, J., Casino, C., and Simmonds, P. (1997). Virus "quasispecies": making a mountain out of a molehill?, Journal of General Virology *78*, 1511-1519.

StatisticsNZ (1997). 1996 New Zealand Census of Population and Dwellings. (Wellington, Statistics New Zealand).

Stern, L. J., Brown, J. H., Jardetzky, T. S., Gorga, J. C., Urban, R. G., Strominger, J. L., and Wiley, D. C. (1994). Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide., Nature *368*, 215-221.

Stuyver, L., Gendt, S., Van Geyt, C., Zoulim, F., Fried, M., Schinazi, R., and Rossau, R. (2000). A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness., Journal of General Virology *81*, 67-74.

Swofford, D. (2000). PAUP*. Phylogenic Analysis Using Parsimony (*and Other Methods). (Massachusetts, Sinaeur Associates, Sunderland).

Takada, S., Kido, H., Fukutomi, A., Mori, T., and Koike, K. (1994). Interaction of hepatitis B virus X protein with a serine protease, typtase TL2, as an inhibitor., Oncogene *9*, 341-348.

Takahashi, K., Aoyama, K., Ohno, N., Iwata, K., Akahane, Y., Baba, K., Yoshizawa, H., and Mishiro, S. (1995). The precore/core promoter mutant ($T^{1762}A^{1764}$) of hepatitis B virus: clinical significance and an easy method for detection., Journal of General Virology *76*, 3159-3164.

Tanaka, K., and Takiguch, M. (2001). Identification of hepatitis B virus - specific CTL epitopes presented by HLA-A*2402, the most common HLA class I allele in East Asia., Journal of Hepatology *34*, 922-929.

The_MHC_sequencing_consortium (1999). Complete sequence and gene map of a human major histocompatibility complex, Nature *401*, 921-923.

Thio, C., Carrington, M., Marti, D., O'Brien, S., Vlahov, D., Nelson, K., Astemborski, J., and Thomas, D. (1999). Class II HLA Alleles and Hepatitis B Virus Persistence in African Americans., The Journal of Infectious Disease *179*, 1004-1006.

Thursz, M., Thomas, H., Greenwood, B., and Hill, A. (1997). Heterozygote advantage for HLA class-II type in hepatitis B virus infection., Nature Genetics *17*, 11-12.

Truant, R., Antunovic, J., Greenblatt, J., Prives, C., and Cromlish, A. (1995). Direct interaction of the hepatitis B virus HBx protein with p53 leads to inhibition by HBx of p53 response element-directed transactivation., Journal of Virology *69*, 1851-1859.

Tsai, S., Chen, M., Yeh, C., Chu, C., Lin, A., Chiou, F., Chang, T., and Liaw, Y. (1996). Purification and Characterisation of a Naturally Processed Hepatitis B Virus Peptide Recognised by CD8+ Cytotoxic T Lymphocytes., Journal of Clinical Investigation *97*, 577-584.

Twu, J., and Schloemer, R. (1987). Transcriptional *trans*-Activating Function of Hepatitis B virus., journal of Virology *61*, 3448-3453.

Uchida, T., Saitoh, T., and Shinzawa, H. (1997). Mutations of the X region of hepatitis B virus and their clinical implications., Pathology International *47*, 183-193.

www.who.int/inf-fs/en/fact204.html (2000). Hepaptitis B (The World Health Organisation).

Xu, Z., Jensen, G., and Yen, T. (1997). Activation of Hepatitis B virus S Promoter by the Viral Large Surface Protein via Induction of Stress in the Endoplasmic Reticulum., Journal of Virology *71*, 7387-7392.

Zhang, Y., Nordenfelt, E., and Hansson, B. (1996). Increasing Heterogeneity of the 'a' Determinant of HBsAg Found in the Presumed Late Phase of Chronic Hepatitis B Virus Infection., Scandinavian Journal of Infectious Diseases *28*, 9-15.

Amendments

These are amendments which were recommended to fulfil the requirements for this thesis. They are presented in the same order as in which they appear in the thesis.

Section 1.1 (page 2) The discussion of the four transcripts should be discussion of the four open reading frames.

Section 2.1 (page 15) the statement "Both clinical profiles have precore mutations terminating HBeAg expression" is not correct. The group stated in the previous paragraph are HBV DNA negative.

Section 2.1 (page 17) the statement "Loss of HBeAg leads to a reduced immune response..." should read "Loss of HBeAg leads to a reduced immune tolerance..."

Section 2.1 (page 17) the statement "the viral encapsidation signal spans both the precore and core genes..." should read "The viral encapsidation signal spans a region (nt 1852-1930) of the precore/core genes."

Section 2.1 (pages 17 and 18) the four major mis-sense mutations that are found in the precore region have an effect on the secondary structure of the RNA not the DNA.

Section 2.1 (page 17) "..loss of HBeAg leads to a reduced immune response, and thus selection against the core protein." Should read "..loss of HBeAg leads to a enhanced immune response, and thus selection against the core protein."

Section 4.1.1 (page 42) In discussion of the clinical groupings the statement "The last two clinical groups (+- and --) are negative for HBeAg, and have a lower viral titre compared to..." should not be so categorical but should only be suggested.

Section 4.1.4 (page 60) The point mutation at position 363 in SA18_1998 results in a shortened protein by 158 amino acids not 330. And the point mutation in WA13_1998 results in 11 amino acids lost from the C-terminus of the protein, not the N-terminus.