ANALYSIS OF HEPATITS B VIRUS DNA INTEGRATED INTO THE GENOMES OF RODENT CELLS

RONALD J. JACKSON

A thesis presented for the Degree of Doctor of Philosophy

University of Edinburgh

September 1987
I hereby declare that I alone have composed this thesis, and that, except where stated, the work presented within is my own.
ABSTRACT

This thesis reports the construction of genomic DNA libraries, using chromosomal DNA from rodent derived cells which contain chromosomally integrated hepatitis B virus DNA sequences. Chromosomal DNA, isolated from the eukaryotic tissue cell lines, mouse L/130·4/TK154 and Rat 2/130·4/TK4, was cloned into the bacteriophage lambda vectors λEMBL4 and λCh34, and screen by plaque hybridisation for the presence of chromosomally integrated hepatitis B virus DNA sequences.

Five recombinant clones were isolated and the integration patterns of the hepatitis B virus DNA analysed. The nucleotide sequences of the viral-cellular DNA junctions were determined. Viral DNA integration sites were observed at hepatitis B virus nucleotide positions 994, 1655 and 2300 ± 100. Only one of the clones contained a complete hepatitis B viral genome, from nucleotide positions 994-3182/1-1460. The viral-cellular junction at position 994 is situated between the preS1 "TATA" box promoter and the putative surface gene promoter. Transfection of mammalian tissue culture cells with above clone resulted in expression of a 2·3 kb poly(A)^+ RNA, which hybridised to radioactively labelled hepatitis B virus DNA. Examination of the culture medium, from the transformed cells by radioimmunoassay, indicated that immunologically reactive HBsAg was expressed. This indicates that the proposed surface gene promoter is transcriptionally active in the absence of the preS1 "TATA" box promoter.

The results from transient transformation assays, for the expression of HBsAg, indicate that the major surface gene promoter is contained between hepatitis B virus nucleotide positions 1098 and 1264. No upstream distal transcriptional regulatory sequences were identified. Therefore, all hepatitis B virus sequences necessary for expression of the surface gene are contained between nucleotide positions 1098 and 84. Nucleotide comparisons of the sequenced mammalian hepatitis B "like" viruses, indicate that, the proposed hepatitis B virus surface promoter nucleotide sequence is highly conserved among the mammalian hepadnaviruses.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title Page</td>
<td>1</td>
</tr>
<tr>
<td>Declaration</td>
<td>ii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>ix</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>x</td>
</tr>
<tr>
<td>Chapter One. Introduction: Molecular Biology of the Hepatitis B virus</td>
<td></td>
</tr>
<tr>
<td>1.1 The Biology of Hepatitis B Infection</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Taxonomy: The Hepadnaviridae</td>
<td>2</td>
</tr>
<tr>
<td>1.3 The Viral Genome</td>
<td>9</td>
</tr>
<tr>
<td>1.4 Open Reading Frames and Viral Proteins</td>
<td>21</td>
</tr>
<tr>
<td>1.4.1 The preS/Surface Gene</td>
<td>21</td>
</tr>
<tr>
<td>1.4.2 The preC/Core Gene</td>
<td>25</td>
</tr>
<tr>
<td>1.4.3 The Polymerase ORF</td>
<td>28</td>
</tr>
<tr>
<td>1.4.4 The X ORF</td>
<td>29</td>
</tr>
<tr>
<td>1.5 Transcription and Viral Gene Expression</td>
<td>30</td>
</tr>
<tr>
<td>1.5.1 The Viral Transcripts</td>
<td>30</td>
</tr>
<tr>
<td>1.5.2 The Polyadenylation Signal</td>
<td>33</td>
</tr>
<tr>
<td>1.5.3 The Surface mRNAs</td>
<td>35</td>
</tr>
<tr>
<td>1.5.4 The Core mRNAs</td>
<td>41</td>
</tr>
<tr>
<td>1.5.5 The HBV Enhancer Element</td>
<td>46</td>
</tr>
<tr>
<td>1.5.6 The X mRNAs</td>
<td>49</td>
</tr>
<tr>
<td>1.6 Viral Replication</td>
<td>51</td>
</tr>
<tr>
<td>1.7 Viral Infection and Hepatocellular Carcinoma</td>
<td>59</td>
</tr>
<tr>
<td>1.8 The Aims of this Thesis</td>
<td>63</td>
</tr>
</tbody>
</table>
Chapter Two. Materials and Methods

2.1 Bacterial Strains 66
2.2 Plasmids and Bacteriophages 67
2.3 Suppliers of Laboratory Reagents 68
2.4 Bacteriological Media 69
2.5 Isolation of Plasmid and Bacteriophage Lambda DNA 70
  2.5.1 Large Scale Isolation of Plasmid DNA 70
  2.5.2 Alkaline-SDS Preparation of Plasmid DNA 71
  2.5.3 Preparation of Bacteriophage Lambda DNA 71
2.6 General Nucleic Acid Manipulation Methods 73
  2.6.1 Phenol Extraction of Nucleic Acid Solutions 73
  2.6.2 Precipitation of Nucleic Acids 73
    a) Ethanol precipitation 73
    b) Isopropanol precipitation 74
  2.6.3 In Vitro Enzymic Reactions 74
    a) Digestion of DNA with restriction endonucleases 74
    b) Modification of recessed 3' termini of DNA 74
    c) In vitro recombination of DNA fragments 75
2.7 Sucrose Gradient Fractionation of DNA Fragments 75
2.8 Gel Electrophoresis of Nucleic Acids 76
  2.8.1 Agarose Gel Electrophoresis of DNA 76
  2.8.2 Gel Photography 76
  2.8.3 Electrophoresis of RNA on Agarose Gels 77
  2.8.4 Electrophoresis of RNA on Polyacrylamide Gels 77
  2.8.5 Staining of Nucleic Acids Separated in Gels Containing Formaldehyde 78
    a) Ethidium bromide staining 78
    b) Silver staining of nucleic acids 78
  2.8.6 Recovery of DNA from Low Melting-Temperature Agarose Gels 78
2.9 Radioactive Labelling of DNA
  2.9.1 Nick Translation
  2.9.2 Random Priming using Hexadeoxyribonucleotide Primers
  2.9.3 5' DNA Terminus Labelling
  2.9.4 Measurement of Radioactivity in Nucleic Acids
    a) Precipitation with trichloroacetic acid
    b) Absorption to DE81 paper
  2.9.5 Sephadex Gel Filtration

2.10 Transformation of Escherichia coli

2.11 In Vitro Packaging of Bacteriophage Lambda DNA

2.12 Nucleic Acid Hybridisation
  2.12.1 Plaque Detection by Blotting
  2.12.2 Transfer of DNA to Membranes (Southern blotting)
  2.12.3 Transfer of RNA to Membranes (Northern blotting)
  2.12.4 DNA Blot Hybridisation
  2.12.5 RNA Blot Hybridisation
  2.12.6 Autoradiography

2.13 Electron Microscopy of DNA Heteroduplexes

2.14 DNA Sequencing
  2.14.1 Preparation of M13 Replicative Forms
  2.14.2 Cloning into the M13 Vectors
  2.14.3 Preparation of Single-Strand Template DNA
  2.14.4 DNA Sequencing Reactions
  2.14.5 Polyacrylamide Gel Electrophoresis

2.15 In Vitro Transcription of Eukaryotic Genes

2.16 Eukaryotic Cell Culture
  2.16.1 Cell Lines and Growth Conditions
  2.16.2 Transformation of Mammalian Cells with Cloned DNA
    a) Calcium phosphate precipitation
    b) Transient assay procedure
  2.16.3 Preparation of Eukaryotic DNA from Tissue Culture Cells
  2.16.4 Preparation of Eukaryotic RNA from Tissue Culture Cells
2.16.5 Isolation of Total Nucleic Acid from Tissue Culture Cells 102
2.16.6 In Vivo Labelling of Proteins 102
2.17 Protein and Immunological Techniques 103
  2.17.1 Determination of Protein Concentration 103
  2.17.2 Radioactive Labelling of Proteins using Iodogen 104
  2.17.3 Separation of Proteins by Polyacrylamide Gel Electrophoresis 105
  2.17.4 Staining of Proteins Separated on Polyacrylamide Gels 106
  2.17.5 Western Blotting 106
  2.17.6 Radioimmunoassays 107
  2.17.7 Double Antibody Immunoprecipitation 107

Chapter Three. Results and Discussion
Cloning and Analysis of Integrated HBV Sequences

3.1 Construction and Screening of Genomic DNA Libraries 109
  3.1.1 Characterisation of Integrated HBV DNA and Cloning Vectors 109
  3.1.2 Construction of the Genomic Libraries 115
3.2 Analysis of the Genomic Clones 117
  3.2.1 λLTK154-H7 118
  3.2.2 λLTK154-R3 123
  3.2.3 λTK4-1 and λTK4-2 130
  3.2.4 λTK4-5 135
3.3 Discussion: Integrated HBV DNA in Hepatocellular Carcinomas and Derived Cell Lines 142
Chapter Four. Results and Discussion

Transcriptional Control of the Major Surface Promoter

4.1 Analysis of HBsAg Expression Directed by Chromosomally Integrated HBV DNA

4.1.1 Transfection of Eukaryotic Cells with XLTK154-H7

4.1.2 Analysis of Integrated DNA in the Transformed Cells

4.1.3 Analysis of HBV mRNA Directed by the Integrated HBV Sequences

4.1.4 HBsAg Expression Directed by Integrated HBV Sequences

4.2 DNA Nucleotide Comparison of the Hepadnaviridae Major Surface Promoters

4.2.1 Nucleotide Comparisons

4.2.2 The Nuclear Factor-1 Binding Site

4.3 E. coli Methylation of HBV Sequences

4.4 In Vitro Transcription Assays

4.5 Transient Assay Expression of HBsAg Using Mouse L Cells

4.5.1 Preliminary Transient Assay Experiments

4.5.2 Transient Assays for the Expression of HBsAg: Transcriptional Regulatory Signals

4.6 Discussion: HBV Sequences Involved in the Regulation of Transcription of the Surface Gene

Appendix I Nucleotide Sequence of HBV130

Bibliography
ACKNOWLEDGEMENTS

I am indebted to Kenneth Murray, my supervisor, and David Finnegan for their guidance throughout the past five years. I also thank the other members of the research group, past and present, for their advice and instruction particularly Sandra Bruce, Heather Houston, John Pugh and Christian Weber. I am grateful to Graham Brown for the photographic work and Pamela Beattie for assistance with the heteroduplex analysis. I am indebted to my wife Valerie for her support and encouragement over the last two years. Finally, I thank the British Council and The Association of Commonwealth Universities for financial support.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisacrylamide</td>
<td>( N,N')-methylene bisacrylamide</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CAT</td>
<td>chloramphenicol acetyltransferase</td>
</tr>
<tr>
<td>cccDNA</td>
<td>covalently closed circular DNA</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CP</td>
<td>core gene promoter</td>
</tr>
<tr>
<td>cpm</td>
<td>counts per minute</td>
</tr>
<tr>
<td>CTF</td>
<td>CCAAT transcription factor</td>
</tr>
<tr>
<td>ddNTP</td>
<td>dideoxynucleotide triphosphate</td>
</tr>
<tr>
<td>DEAE</td>
<td>diethylaminoethyl</td>
</tr>
<tr>
<td>Dex</td>
<td>dexamethasone</td>
</tr>
<tr>
<td>( d\text{H}_2\text{O} )</td>
<td>distilled ( \text{H}_2\text{O} )</td>
</tr>
<tr>
<td>DHBV</td>
<td>duck hepatitis B virus</td>
</tr>
<tr>
<td>DME</td>
<td>Dulbecco's modified Eagles' medium</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>DR1</td>
<td>direct repeat one</td>
</tr>
<tr>
<td>DR2</td>
<td>direct repeat two</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetra-acetate</td>
</tr>
<tr>
<td>En</td>
<td>Enhancer</td>
</tr>
<tr>
<td>FCS</td>
<td>foetal calf serum</td>
</tr>
<tr>
<td>FTL</td>
<td>freeze thaw lysate</td>
</tr>
<tr>
<td>GA-HSA</td>
<td>glutaraldehyde-polymerised human serum albumin</td>
</tr>
<tr>
<td>gp28</td>
<td>28,000 dalton glycoprotein</td>
</tr>
<tr>
<td>GSHV</td>
<td>ground squirrel hepatitis virus</td>
</tr>
<tr>
<td>HBcAg</td>
<td>hepatitis B core antigen</td>
</tr>
<tr>
<td>HBeAg</td>
<td>hepatitis B e antigen</td>
</tr>
<tr>
<td>HBsAg</td>
<td>hepatitis B surface antigen</td>
</tr>
<tr>
<td>HBV</td>
<td>hepatitis B virus</td>
</tr>
<tr>
<td>HBxAg</td>
<td>hepatitis B X antigen</td>
</tr>
<tr>
<td>HCC</td>
<td>hepatocellular carcinoma</td>
</tr>
<tr>
<td>HEPES</td>
<td>2-hydroxyethylpiperazine-( N'2 )-ethanesulphonic acid</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-thio-( \beta )-D-galactoside</td>
</tr>
<tr>
<td>kb</td>
<td>kilo base</td>
</tr>
<tr>
<td>kd</td>
<td>kilo dalton</td>
</tr>
<tr>
<td>KHV</td>
<td>kangaroo hepatitis virus</td>
</tr>
<tr>
<td>MEM</td>
<td>minimal essential medium</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-(( N )-morpholino)propanesulphonic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>NF-1</td>
<td>nuclear factor 1</td>
</tr>
<tr>
<td>ocDNA</td>
<td>open circular DNA</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>p25</td>
<td>25,000 dalton protein</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate-buffered saline</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>pfu</td>
<td>plaque forming unit</td>
</tr>
<tr>
<td>pHSA</td>
<td>polymerised human serum albumin</td>
</tr>
<tr>
<td>[r( r )]</td>
<td>small terminal redundancy on HBV minus strand</td>
</tr>
<tr>
<td>[R( R )]</td>
<td>terminal redundancy on HBV pregenomic RNA</td>
</tr>
<tr>
<td>RF</td>
<td>replicative form</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecylsulphate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecylsulphate-polyacrylamide gel</td>
</tr>
<tr>
<td></td>
<td>electrophoresis</td>
</tr>
<tr>
<td>SE</td>
<td>sonicated extract</td>
</tr>
<tr>
<td>SPI</td>
<td>preS1 gene promoter</td>
</tr>
<tr>
<td>SPII</td>
<td>preS2/surface gene promoter</td>
</tr>
<tr>
<td>SSC</td>
<td>standard saline citrate</td>
</tr>
<tr>
<td>SSHV</td>
<td>stink snake hepatitis virus</td>
</tr>
<tr>
<td>SSPE</td>
<td>standard saline phosphate EDTA</td>
</tr>
<tr>
<td>TCA</td>
<td>trichloroacetic acid</td>
</tr>
<tr>
<td>TEMED</td>
<td>$N,N',N,N'$-tetramethylethylenediamine</td>
</tr>
<tr>
<td>THBV</td>
<td>tree squirrel hepatitis B virus</td>
</tr>
<tr>
<td>$T_m$</td>
<td>melting temperature</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>WHV</td>
<td>woodchuck hepatitis virus</td>
</tr>
<tr>
<td>XG</td>
<td>5-bromo-4-chloro-3-indol-β-D-galactoside</td>
</tr>
<tr>
<td>XP</td>
<td>X gene promoter</td>
</tr>
</tbody>
</table>
1.1 The Biology of Hepatitis B Infection

Viral hepatitis B is a contagious infection of man. It has been estimated that there are over 200 million carriers of hepatitis B worldwide providing a large reservoir of infection (Szmuness, 1978). The disease can be transmitted parenterally, sexually and by inoculation with contaminated medical instruments or by blood transfusion (hence the term "serum hepatitis").

The host range of the hepatitis B virus (HBV) is narrow; to date, productive experimental infections have been established only in humans and higher primates (Hirschman et al., 1969; Maynard et al., 1972; Barker et al., 1973, 1975). The virus shows a strong liver tropism, however, recently viral DNA sequences have been detected in low copy number in cells other than hepatocytes (Blum et al., 1983), most commonly associated with blood leukocytes and bone marrow (Romet-Lemonne et al., 1983a, 1983b; Elfassi et al., 1984; Laure et al., 1985; Korba et al., 1986; Noonan et al., 1986). These observations indicate that the tissue tropism of the virus may be broader than once thought.

Primary HBV infections in adults are most often acute, a moderately severe illness characterised by hepatocellular injury and inflammation. Many individuals experience mild or no liver injury despite extensive hepatic infection (Redeker, 1975). However, some patients develop fulminant disease with extensive hepatic necrosis and high mortality (Peters, 1975). Most primary infections are usually self-limited and are resolved by the immune system. The host response to infection results in the development of antibodies to the virus associated antigens HBsAg,
HBcAg and HBeAg (see below) resulting in the complete clearance of the virus and the development of lasting immunity to reinfection (fig. 1.1a).

However, 5%-10% of immunologically normal adults (Redeker, 1975; Hoofnagle et al., 1978) and a much higher proportion of newborn infants (Schweitzer et al., 1973; Tong et al., 1981) do not resolve primary infections, but develop a persistent hepatic infection which is characterised by the persistence of HBsAg in the peripheral blood (fig. 1.1b). As with primary infections, such individuals may be asymptomatic or experience varying grades of chronic liver injury, with a significant number of patients having severe infections that progress to cirrhosis (Redeker, 1975; Peters, 1975). Long-standing carriers, mostly with cirrhosis, have a predisposition to the development of hepatocellular carcinoma (HCC); the incidence of HCC among long-term chronic HBV carriers is more than 100-fold that of age matched noncarriers (Beasley et al., 1981).

1.2 Taxonomy: The Hepadnaviridae

The human hepatitis B virus was considered for many years to be a unique virus. The identification of the aetiological agent responsible for serum hepatitis (viral hepatitis B) followed the discovery of the so-called "Australia antigen" (hepatitis B surface antigen) (Blumberg et al., 1965), the association of this antigen with serum hepatitis (Blumberg et al., 1968; Prince, 1968) and the subsequent discovery of the Dane particle (HBV virion) (Dane et al., 1970). Since these early observations it has become apparent that viruses with similar ultrastructural, molecular and biological features infect other species of animals.

Robinson, (1980) devised the name Hepadnaviridae (hepatitis-DNA-virus) to describe the closely related family of viruses of which HBV is the prototype. The Hepadnaviridae at present consists of possibly seven viruses: four fully characterised viruses, the human hepatitis B virus (HBV), the woodchuck hepatitis virus (WHV) (Summers et al., 1978), the
Figure 1.1 Serology of HBV infection. The appearance of HBV antigens, and antibodies to them, in serum during a) acute and b) chronic infection. There are wide variations in the time at which the various serological markers are found in individual infections; this is indicated in part by the broken lines. Figure adapted from Murray, (1987).
ground squirrel hepatitis virus (GSHV) (Marion et al., 1980a) and the
duck hepatitis B virus (DHBV) (Mason et al., 1980); the partially
classified tree squirrel hepatitis B virus (THBV) (Feitelson et al.,
1986a, 1986b) and two other possible uncharacterised candidates, the stink
snake hepatitis virus (SSHV) and the kangaroo hepatitis virus (KHV)
(Howard, 1986). All the viruses show a strong liver tropism and limited
host range, usually infecting only one species of animal, and share a
common pattern of associated diseases. The mammalian members of the
family also share a limited degree of immunological cross-reactivity
between their major antigenic determinants (Werner et al., 1979; Gerlich
et al. 1980; Cote and Gerin, 1983; Feitelson et al., 1986a, 1986b).

However, Mason et al., (1980) could not detect any serological cross-
reactivity between the viral surface antigens of the duck hepatitis
virus and the mammalian hepatitis viruses. Table 1.1 lists the specific
characteristics of the four characterised hepadnaviruses. The human
hepatitis B virus was the first to be discovered and is considered to be
the type virus of the Hepadnaviridae, the general features of this virus
are described below.

The HBV virion (Dane particle) (fig. 1.2a) is a spherical enveloped virus
with an overall diameter of approximately 42 nm, with an inner electron
dense spherical core, of 27 nm diameter (Dane et al., 1970). The lipid,
protein and carbohydrate containing envelope (Burrell et al., 1973;
Steiner et al., 1974; Gavilanes et al., 1982) bears the hepatitis B
surface antigen (HBsAg) (Dane et al., 1970; Almeida et al., 1971). The
inner core particle or nucleocapsid, which can be released from the
envelope by detergent treatment, bears the hepatitis B core antigen
(HBcAg) (Almeida et al., 1971).

Contained within the core particle is the viral genome, which consists of
a small relaxed circular DNA molecule that is partially single stranded
(Robinson et al., 1974; Summers et al., 1975; Hruska et al., 1977; Landers
et al., 1977). Also contained within the core particles is a DNA
polymerase activity which is capable of "filling in" the single stranded
region of the genome (Hirchman et al., 1971; Kaplan et al., 1973; Robinson
and Greenman, 1974).
<table>
<thead>
<tr>
<th></th>
<th>HBV</th>
<th>WHV</th>
<th>GSHV</th>
<th>DHBV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Virion</strong></td>
<td>Enveloped 42nm spherical with HBsAg 27nm core with HBCAg</td>
<td>Enveloped 45nm spherical with WHsAg 27nm core with WHCAg</td>
<td>Enveloped 47nm spherical with GSHsAg ca. 30nm core with GSHCAg</td>
<td>Enveloped 40-45nm spherical with DHBsAg 27nm core (spikes) with DHBcAg</td>
</tr>
<tr>
<td><strong>Protein kinase</strong></td>
<td>DNA polymerase</td>
<td>DNA polymerase</td>
<td>DNA polymerase</td>
<td>DNA polymerase</td>
</tr>
<tr>
<td><strong>Genome</strong></td>
<td>DNA Circular Single-stranded gap</td>
<td>DNA Circular Single-stranded gap</td>
<td>DNA Circular Most molecules fully double stranded Cohesive ends</td>
<td>DNA Circular Most molecules fully double stranded Cohesive ends</td>
</tr>
<tr>
<td><strong>Surface proteins</strong></td>
<td>HBsAg particles 22nm spherical and filamentous forms</td>
<td>WHsAg particles 20-25nm spherical and filamentous forms</td>
<td>GSHsAg particles 15-25nm spherical and filamentous forms</td>
<td>DHBsAg particles 40-60nm spherical and convoluted forms</td>
</tr>
<tr>
<td><strong>Surface core proteins (kDa)</strong></td>
<td>S:p25, gp28 preS2:gp33, gp36 preS1:gp39, gp42</td>
<td>S:p23, gp27 preS2:gp33, gp36 preS1:gp45, gp47</td>
<td>S:p23, gp27 preS2:gp33, gp36 preS1:gp45, gp47</td>
<td>S:p17,5 preS1:gp36, gp34</td>
</tr>
<tr>
<td><strong>Natural host</strong></td>
<td>Human eastern woodchuck (Marmota monax)</td>
<td>Beechey ground squirrel (Spermophilus beecheyi)</td>
<td>Pekin duck (Anas domesticus)</td>
<td></td>
</tr>
</tbody>
</table>

† Computed prediction.

Examination of electron micrographs of fully double stranded DNA molecules isolated from Dane particles, indicate that the HBV genome is approximately 3200 base pairs in length (Hruska et al., 1977). In addition to the DNA polymerase activity, there is also a protein kinase activity which phosphorylates the viral core antigen protein (Albin and Robinson, 1980). It is presently unclear whether the kinase activity is intrinsic to the core antigen protein or is due to an associated protein encoded by the virus or derived from the host cell.

Numerous particles are found in the serum of HBV infected patients in many fold excess over the virions, these particles bear HBsAg determinants (HBsAg particles). The HBsAg particles have two forms, small spherical particles of 22 nm diameter and filamentous particles 22 nm wide with a variable length of up to several hundred nm (fig. 1.2b). The particles have the same biochemical composition as the virion envelope being composed of lipid, protein and carbohydrate, however, they are devoid of HBCAg and viral DNA and are considered to be empty viral envelope material (Bayer et al., 1968; Dane et al., 1970; Almeida et al., 1971). The reason for the presence of these particles is uncertain, one possibility is that these empty envelope particles may serve to adsorb neutralising surface antibodies during the progression of infection. The HBsAg present in the viral envelope and the HBsAg particles has a major group specific antigenic determinant (a) and two sets of generally mutually exclusive subtype determinants (d/y) and (r/w), giving rise to the four major antigenic subtypes adr, adw, ayr and ayw (LeBouvier, 1971; Bancroft et al., 1972). Additional antigenic variation of the w determinant and other minor determinants add to the complexity of HBsAg (Courouce-Pauty et al. 1976). Wands et al., (1986) have identified isolates of variant human hepatitis B viruses that are immunologically distinct from the known subtypes of HBV with respect to their major antigens, however, they do share antigenic epitopes with HBV that are recognised by monoclonal antibodies. The hepatitis B virus variants also share complementary DNA sequences to HBV. The variants and HBV were shown not to be cross-protective upon challenge in chimpanzees, indicating that their major antigenic determinants are immunologically distinct.
Figure 1.2 A schematic representation of the viral particles and components (not to scale). (a) The HBV virion (Dane particle) consists of an 42 nm diameter envelope carrying HBsAg and a 27 nm diameter nucleocapsid carrying HBCAg (and cryptic HBeAg). The nucleocapsid contains a circular DNA molecule and virion associated enzymic activities. (b) The HBsAg particles, 22 nm diameter spheres and filaments, are empty envelope material carrying HBsAg. (c) The Dane particle open circular DNA (3182 nucleotides) showing the single stranded region, 5' protein and oligoribonucleotide.
DNA

- HBsAg
- HBcAg
- HBeAg (cryptic)
- Polymerase
- Protein Kinase

(a) 27 nm 42 nm

(b) filament

(c) 5' RNA

- 5' Protein

1/3182 2882 (5')

(3') 3100 (5')

NICK
A third antigenic specificity associated with HBV infection is the hepatitis B e antigen (HBeAg) which is found in the serum of infected patients (Magnius and Espmark, 1972). The HBeAg may exist as a free polypeptide or in association with IgG molecules (Tedder and Bull, 1979; Katz et al., 1980; Blanchy et al., 1980). The HBeAg specificity is also associated with the virion nucleocapsid after disruption of core particles by detergent treatment (Ohori et al., 1979; Takahashi et al., 1979). At least two distinct HBeAg (HBeAg/1 and HBeAg/2) epitopes have been localized on the HbcAg polypeptide, which are exposed when the HbcAg protein is unfolded (Budkowska et al., 1979; Ohori et al., 1980, 1984; Ferns and Tedder, 1984).

The Hepadnaviruses have been reviewed by, Summers, (1981); Marion and Robinson, (1983); Tiollais et al., (1981, 1985) and Howard, (1986).

1.3 The Viral Genome

The HBV virion DNA molecule is partially double stranded, thus the DNA molecule contains two strands of unequal length, a complete long strand which is complementary to the viral mRNAs and by convention is designated to be of minus polarity; accordingly, the shorter complementary strand is termed the plus strand. The long strand is not a covalently closed circle, a "nick" exists at a unique site approximately 300 bp from the 5' end of the short strand (Summers et al., 1975; Sattler and Robinson, 1979; Siddiqui et al., 1979) (fig. 1.2c). The circular DNA can be converted to a linear form with single strand cohesive ends by heating under the appropriate conditions and selectively denaturing the ca. 300 bp between the 5' ends of the two strands (Sattler and Robinson, 1979). The circular conformation of the molecule is maintained by hybridisation of cohesive/complementary ends at the 5' termini of each strand. The ends of the HBV virion DNA strands have recently been mapped on the HBV sequence; the 5' end of the minus strand is positioned at nucleotide 3108 with the 3' end probably at position 3100 ± 1. Thus the minus strand contains a short terminal
redundancy of ca. 9 nucleotides at its 5' and 3' ends. The 5' end of the DNA plus strand has been mapped to position 2882 ± 1, with the 3' end at a variable position (see below) (Will et al., 1987). Therefore the cohesive overlap is approximately 220 nucleotides in length from nucleotides 2882-3001.

The single stranded gap present on the HBV virion genome has a highly favoured minimum length of 650 to 700 nucleotides with all the molecules being double stranded upstream (on the plus strand) of position 1420 on the HBV map. Therefore the 3' end of the plus strand is at or downstream of position 1420 with a favoured 3' end between positions 2180 to 2230 (Delius et al., 1983). By contrast, on the genome of the duck hepatitis B virus the region of the single strand gap is smaller, on average, with many fully double-stranded molecules present in DHBV virions (Mason et al., 1981; Summers and Mason, 1982).

The 5' ends of both strands of DNA appear to be "protected" in a manner that prevents phosphorylation by polynucleotide kinase (Gerlich and Robinson, 1980). It has been shown that a protein is attached to the 5' end of the long strand; the attachment is presumably covalent since heating at 90°C in detergent or treatment with strong alkali failed to remove the attached protein (Gerlich and Robinson, 1980). The chemical nature to the blockage to phosphorylation of the 5' end of the short strand until recently remained an enigma. Lein et al., (1986) have demonstrated that the 5' termini of the short strand of the DHBV genome contain a capped oligoribonucleotide of 18 or 19 bases in length, with a possible cap structure (m7GpppAA). Oligoribonucleotides have also been demonstrated to be attached to the 5' ends of the short strands of the woodchuck hepatitis virus, the ground squirrel hepatitis virus and the human hepatitis B virus (Seeger et al.,1986; Will et al. 1987). If a similar capped structure is also present on the oligoribonucleotide attached to the 5' terminus of the HBV short strand, this would prevent phosphorylation by polynucleotide kinase, as observed by Gerlich and Robinson, (1980). The significance of the protein and RNA covalently bound to the 5' ends of the DNA strands is presented below (see section, 1.6 Viral Replication).
In the presence of nucleoside triphosphates the product of the endogenously primed DNA polymerase reaction is a fully double stranded DNA molecule (Kaplan, et al., 1973; Robinson and Greenman, 1974; Summers et al., 1975; Hruska et al., 1977; Landers et al., 1977). These double stranded DNA molecules have been cloned into bacterial vectors and their nucleotide sequences determined. The sequenced Hepadnaviruses are listed in table 1.2.

Many different nomenclature systems have been used for the Hepadnaviridae. A common system used is that of using a unique Eco R1 restriction endonuclease site as position one; this site is present in most isolates of HBV (Galibert et al., 1979). However, since this restriction site is not present in some isolates of HBV (Burrell et al., 1979; Gough and Murray, 1982) and does not correspond to a uniform origin in the other animal viruses, the numbering system of Pasek et al., (1979) has been adopted here. The numbering system of all the hepadnaviruses have been changed such that position one corresponds to the start of the coding sequence for the "mature" HBcAg polypeptide, starting at the predicted triplet amino acid translation Met-Asp-Ile common to all the viruses. This corresponds to the second ATG methionine codon of the PreC/C open reading frame for HBV, GSHV and DHBV, or the third initiation codon for WHV which corresponds to the second ATG codon of HBV and GSHV. The representative genetic maps of each type of virus is presented in fig. 1.3 (HBV), fig. 1.4 (WHV), fig. 1.5 (GSHV) and fig. 1.6 (DHBV).

The genetic organisation of the Hepadnavirus genomes are similar. Analysis of the complete nucleotide sequence of all Hepadnaviruses examined shows that only the major open reading frames of the long minus strand are conserved, while major reading frames found on the plus strand are not conserved among the Hepadnaviridae. Four major open reading frames are found in all sequenced mammalian Hepadnaviruses, these are termed PreS/Surface, PreC/Core, Polymerase and X (or B). Only three open reading frames are found on the DHBV sequence corresponding to PreS/Surface, PreC/Core and Polymerase.
Table 1.2 List of sequenced Hepadnaviruses.

<table>
<thead>
<tr>
<th>Hepadnavirus</th>
<th>Nucleotide length</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBVadyw (HBV130)</td>
<td>3182 bp</td>
<td>Pugh et al., (1986)</td>
</tr>
<tr>
<td>HBVadyw (HBV14)</td>
<td>3188 bp</td>
<td>Will et al., (1982)</td>
</tr>
<tr>
<td>HBVadr</td>
<td>3188 bp</td>
<td>Ono et al., (1983)</td>
</tr>
<tr>
<td>HBVadr(2)</td>
<td>3215 bp</td>
<td>Kobayashi and Koike, (1984)</td>
</tr>
<tr>
<td>HBVadr4</td>
<td>3214 bp</td>
<td>Fujiiyama et al., (1983)</td>
</tr>
<tr>
<td>HBVadw</td>
<td>3200 bp</td>
<td>Ono et al., (1983)</td>
</tr>
<tr>
<td>HBVadw2</td>
<td>3221 bp</td>
<td>Valenzuela et al., (1980)</td>
</tr>
<tr>
<td>HBVayr</td>
<td>3215 bp</td>
<td>Okamota et al., (1986)</td>
</tr>
<tr>
<td>HBVayw</td>
<td>3182 bp</td>
<td>Galibert et al., (1979)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bichko et al., (1985)</td>
</tr>
<tr>
<td>DHBV</td>
<td>3021 bp</td>
<td>Mandart et al., (1984)</td>
</tr>
<tr>
<td>GSHV</td>
<td>3311 bp</td>
<td>Seeger et al., (1984)</td>
</tr>
<tr>
<td>WHV</td>
<td>3308 bp</td>
<td>Galibert et al., (1982)</td>
</tr>
<tr>
<td>WHV2</td>
<td>3320 bp</td>
<td>Kodama et al., (1985)</td>
</tr>
</tbody>
</table>
Figure 1.3 Genetic map of the human hepatitis B virus.

The map shows the positions of the major open reading frames, the mapped transcriptional initiation sites and associated transcriptional regulatory signals. Also shown are the positions of the nuclear protein binding sites (U1 to U3, E and NF-1) and the direct repeats (DR1 and DR2). Abbreviations: CP (Core promoter), EN/XP (Enhancer element and X promoter), SPI (preS1 gene promoter) and SPII (preS2 and Surface gene promoter). References: Stibbe and Gerlich, (1983); Heermann et al., (1984); Shaul et al., (1985,1986a); Tognoni et al., (1985); Pugh et al., (1986); Shaul and Ben-Levy, (1987); Treinin and Laub, (1987) and Yaginuma et al., (1987b).
HBV

- Core Promoter Region
  - Pol end 2902
  - DR1 3106-3116
  - 5' end Core/Pregenome mRNA 3099
  - preC start 3096
  - 5' end preC mRNA 3063, 3072

- Enhancer/X Promoter Region
  - X start 2656
  - 5' ends X mRNA 2590, 2612, 2620
  - E Binding Site 2460
  - UE1 Binding Site 2309
  - UE2 Binding Site 2252
  - UE3 Binding Site 2161
  - X ORF
  - X end 3117

- preC
  - 5' end preC mRNA 3063, 3072
  - preC start
  - 5' end preC mRNA

- Core
  - Core start
  - 16 polyadenylation signal (TATAAA)
  - 5' end Core mRNA
  - preS1 Promoter (TATATAAA)
  - 5' end preS1 mRNA
  - 907 5' end preS1 mRNA
  - 948 preS1 start
  - 876 preS1 Promoter (TATATAAA)
  - 1068 NF-1 Binding Site
  - 1256 5' end preS2 mRNA
  - 1276 5' end Surface mRNA
  - 1280 (EcoR1)
  - 1287 Alternative 5' end Surface mRNA
  - 1437 Surface start

- Surface
  - Surface end 2114
  - Surface/preS2 Promoter
  - 1256 5' end preS2 mRNA
  - 1272 preS2 start

- Polymerase
  - 5' end preS2 mRNA
  - 1272 preS2 start
  - 1276 5' end Surface mRNA
  - 1280 (EcoR1)
  - 1287 Alternative 5' end Surface mRNA
  - 1437 Surface start
Figure 1.4 Genetic map of the woodchuck hepatitis virus.

The map shows the positions of the major open reading frames, the mapped transcriptional initiation sites and associated transcriptional regulatory signals. Also shown are the positions of the direct repeats (DR1 and DR2). Abbreviations: CP (Core gene promoter), E (enhancer) and SP (Surface gene promoter). References: Galibert et al., (1982); Moroy et al., (1985) and Schaeffer et al., (1986).
5' end Core/Pregenome mRNA 3213, 3238 X end 3213
preC start 3198, 3219
Core promoter region
Pol end 3043
DR1 3230-3239
5' end minor mRNA 2996
3308/1 Core start
16 polyadenylation signal (TATAAA)
DR2 3008-3017
5' end minor mRNA 2996
Pol start
564 Core end
407 Pol start
972 preS1 start
1267 5' end minor (preS2) mRNA
1288 EcoR1
Surface promoter
1404 preS2 start
1433 5' end Surface mRNA
1446 Alternative 5' end Surface mRNA
1584 Surface start

Homology to HBV Enhancer

Surface end 2249
Surface promoter
1404 preS2 start
1433 5' end Surface mRNA
1446 Alternative 5' end Surface mRNA
1584 Surface start

Surface end 2249
Surface promoter
1404 preS2 start
1433 5' end Surface mRNA
1446 Alternative 5' end Surface mRNA
1584 Surface start

Homology to HBV Enhancer

X start 2791
X ORF
preC
Core
Polymerase

WHV
Figure 1.5 Genetic map of the ground squirrel hepatitis virus

The map shows the positions of the major open reading frames, the mapped transcriptional initiation sites and associated transcriptional regulatory signals. Also shown are the positions of the direct repeats (DR1 and DR2). Abbreviations: CP (Core gene promoter), E (Enhancer) and SP (Surface gene promoter). References: Seeger et al., (1984); Enders et al., (1985) and Schaeffer et al., (1986).
5' end Core/Pregenome mRNA 3227
preC start 3222
X end 3207
5' end preC mRNAs 3198, 3205
Core promoter region
Pol end 3046
DR2 3010-3020
3311/1 Core start
16 polyadenylation signal (TATAAA)
404 Pol start
561 Core end
882 EcoR1
969 preS1 start
Surface promoter
1396 5' end preS2 mRNA
1407 preS2 start
1439 5' end Surface mRNA
1587 Surface start
Figure 1.6 Genetic map of the duck hepatitis B virus

The map shows the positions of the major open reading frames, the mapped transcriptional initiation sites and associated transcriptional regulatory signals. Also shown are the positions of the direct repeats (DR1 and DR2). Abbreviations: CP (Core gene promoter), SPI (preS gene promoter) and SPII (Surface gene promoter). References: Mandart et al., (1984); Buscher et al., (1985) and Pugh et al., (1987).
DR1 2910-2921
5' end Core/Pregenome mRNA 2905
Pol end 2902
Core promoter region
preC start 2893
DR2 2852-2873
3021/1 Core start
126 polyadenylation signal (AATAAA)
EcoR1

preC
Core
preS
Surface
Polymerase

preS promoter region
1107, 1115 5' ends preS mRNAs
1173 preS start
1360 5' end Surface mRNA
786 Core end
371 EcoR1
395 Pol start
Surface end 2159
Surface start 1659

The genetic organisation of the human hepatitis B virus used in this discussion is based upon the nucleotide sequence of the HBV genomic DNA cloned in the plasmid pHBV130 (Pasek et al., 1979; Gough and Murray, 1982; Pugh et al., 1986) (see Appendix I).

1.4 Open Reading Frames and Viral Proteins

1.4.1 The PreS/Surface Gene

The preS/Surface open reading frame (ORF), position 948-2114, has a total coding capacity for a ca. 42,800 dalton protein. The major HBsAg proteins isolated from the viral envelope or HBsAg 22 nm particles have apparent molecular weights of 25,000 (p25) and 28,000 (gp28) daltons (Peterson et al., 1977). Together with the major envelope proteins a variable number of minor component proteins with apparent sizes of up to 100,000 daltons have been repeatedly observed in HBsAg preparations (reviewed by Robinson, 1977). The proteins observed to be greater than 42,000 daltons are now considered to be aggregates of the HBsAg proteins (Koistinen, 1980; Heermann et al., 1984). The assignment of the p25/gp28 proteins to the PreS/Surface ORF was possible by comparison of the known N-terminal amino acid sequence of the HBsAg protein (Peterson et al. 1977), with the predicted amino acid translation of the HBV sequence. This positioned the major surface gene starting at the ATG methionine codon at position 1437 on the HBV sequence and terminating at 2114, giving a coding capacity for a 226 amino acid protein of molecular weight ca. 25,400 daltons.

The surface ORF is preceded by an in-frame ORF with a coding capacity for an extra 163 amino acids. This region is termed the preS region (948-1436). The preS region is further subdivided into the preS1 (948-1271) and preS2 (1272-1436) regions (fig 1.3). Recently it has been shown that proteins are translated from the third ATG (preS2) (or fourth, depending upon the subtype) and the first ATG (preS1) methionine initiation codons of the preS region, and co-terminating with the major
surface proteins. Therefore, the preS2 and preS1 proteins contain the entire amino acid sequence of the major surface p25 protein with an additional 55 and 163 amino acids respectively. The preS translated products have apparent molecular weights of 33,000 (gp33) and 36,000 (gp36) daltons for the preS2 proteins, and 39,000 (p39) and 42,000 (gp42) daltons for the preS1 proteins (Stibbe and Gerlich, 1983; Heermann et al., 1984).

The difference in apparent molecular weight of the major HBsAg proteins p25/gp28 is due to the gp28 protein being glycosylated between residues 121 and 170, most probably a N-linked glycan on the asparagine residue at position 146 (Peterson et al., 1977; Peterson, 1981). The preS2 glycoproteins gp33/gp36 both have a mannose rich glycan at asparagine residue 4 on their amino acid sequence, 51 amino acid positions upstream from the start of p25/gp28. The difference in molecular weight of the preS2 proteins is due to an extra glycan group on gp36 in an analogous position to gp28 (Stibbe and Gerlich, 1983). Again the size difference between the preS1 proteins p39 and gp42 is due to the glycan group bound to the surface protein part of the molecule. The mannose rich glycan of the preS2 proteins is not present on the preS1 proteins (Heermann et al., 1984). Persing et al., (1987) have recently shown that the preS1 proteins are acylated, with myristic acid covalently bound to the amino terminus. The amino acid motif, methionine-glycine, is found at the amino terminus of the predicted amino acid sequences of the preS1 proteins of all the Hepadna viruses. It is proposed that following cleavage of the methionine residue, an amide linkage is formed between the COOH group of myristic acid and the exposed amino group of glycine (Persing et al. 1987).

The different subtypes of HBV vary in their coding capacity for the preS1 derived polypeptides. The sequenced subtype HBVayw (Galibert et al., 1979) like the HBV sequence cloned in pHBV130, has a preS region of a 163 codons, while most of the remaining sequenced subtypes have an additional 11 codons at the amino termini of the preS1 region. The insertion or deletion resulting in this difference of coding capacity is bounded by a 7 base direct repeat of GCATGGG containing the respective
methionine initiation codon of the preS1 region. It is possible that this evolutionary difference could result from insertion or deletion by homologous recombination via the direct repeat. It should be noted that the sequenced subtype HBV adw (Ono et al., 1983) while having the same preS1 coding capacity as HBV ayw, does however contain DNA sequences homologous to the extra sequences found in the other subtypes. The furthest upstream ATG initiation codon is absent with the additional sequences being only 12 nucleotides in length as opposed to 33 nucleotides in the other subtypes. In agreement with the nucleotide sequence data, p39 and gp42 from isolates of HBV adw, presumably with the same coding capacity for preS1 as HBV adw2 (Valenzuela et al., 1980), are consistently larger than those of HBV ayw, with a difference in size of ca. 1000-1500 daltons in apparent molecular weight (Heermann et al., 1984). This indicates that the coding sequence of p39/gp42 starts with the first AUG codon of the preS/Surface ORF, with the first methionine codon of the HBV ayw and HBV adw preS1 genes corresponding to the second in-frame methionine of the other subtypes.

The three morphological forms of the HBsAg; the Dane particle, the 22 nm spheres and filaments all contain the above six proteins, however, in very different ratios. Heermann et al. (1984) have demonstrated that upon silver staining the relative abundance of the proteins differ depending upon the morphological form. The relative staining intensities suggest that the virion particle (and filaments to a lesser extent) may contain up to 20 times more p39 and gp42, as do the 22 nm HBsAg particles. Recent reports indicate that the preS1 proteins are not secreted from the cell, and when artificially over-expressed, inhibit the secretion of the smaller preS2 and major surface proteins (Cheng et al., 1986; Chisari et al., 1986; Persing et al., 1986; Standring et al., 1986; McLachlan et al., 1987; Ou and Rutter, 1987). The preS1 proteins are only efficiently exported from cells as part of the virion envelope. These observations are consistent with the role for the preS1 proteins in virion assembly. The N-terminal myristic acid (Persing et al., 1987) could be important in anchoring the N-terminus of the preS1 protein in cellular membranes. It has been observed that the preS1 proteins accumulate in the Golgi complex and are not secreted (Ou and Rutter,
1987). In contrast, the preS2 and major HBsAg proteins are associated with the endoplasmic reticulum, from where they are transported to the Golgi complex and secreted (Patzer et al., 1986; Ou and Rutter, 1987).

The presence of the accumulated preS1 protein in the Golgi complex could cause the smaller preS2 and surface proteins to accumulate and therefore, inhibit their secretion. The preS1 imposed restriction on the transport of HBsAg particles may result from intermolecular association of the proteins inhibiting secretion of the 22 nm particles, thus favouring the formation of the larger filaments and the HBV virion in association with the core particle (Ou and Rutter, 1987). The distribution of the surface associated proteins in the 22 nm particles, filaments and virion, is consistent with the above hypothesis. The expression of the preS1 protein is regulated by separate promoter (see section, 1.5.3 The Surface mRNAs), and it is conceivable that expression of the preS1 and other HBsAg proteins is independently regulated, resulting in phasing the formation of the different structures during the viral life cycle.

Binding sites for polymerised human serum albumin (pHSA) have been associated with the preS2 protein present in the HBV virion (Machida et al., 1983,1984). It is postulated that the attachment of HBV to hepatocytes is mediated by species specific pHSA-receptors, present on both the cells and the virion, to pHSA (Imai et al., 1979; Alberti et al., 1982). However, the putative role of the pHSA receptors in pathogenesis and disease remains controversial. The postulate is based on the observation that preS2 proteins bind glutaraldehyde-polymerised HSA (GA-HSA) in vitro. However, naturally occurring pHSA, unlike GA-HSA, does not bind significantly to the HBsAg proteins (Yu et al., 1985) therefore casting doubt on the above hypothesis. Although, the possibility that pHSAs do mediate viral attachment in vivo can not be totally excluded. Neurath et al., (1986) have observed that a synthetic peptide corresponding to preS1 amino acid residues 10 to 36 (or 21 to 47 for the other subtypes) is recognised by both hepatocyte receptors and anti-HBV antibodies. Neurath et al., (1986) propose that the amino acid
sequence present on the preS1 protein may facilitate attachment to hepatocytes.

1.4.2 The PreC/Core Gene

The PreC/Core ORF (3096-3182/1-549) has a coding capacity for a 212 amino acid protein of molecular weight ca. 24,300 daltons. The core ORF was originally defined by early cloning and expression studies conducted in *Escherichia coli* (Burrell *et al.*, 1979; Pasek *et al.*, 1979). Immunologically active HBcAg protein was produced in *E. coli* and was found to be translated from the methionine codon at position one. In this HBV isolate, the initiation codon was preceded by an in-frame stop codon six nucleotides upstream to the initiation codon (Pasek *et al.*, 1979). Therefore, the core gene (1-549) as defined, could code for a polypeptide of 183 amino acids with a molecular weight of ca. 21,000 daltons.

Analysis of other sequenced HBV genomes, as well as that contained in pHBV130, indicates that the core gene is usually preceded by an in-frame 29 codons. The isolate used by Burrell *et al.*, (1979) and Pasek *et al.*, (1979) was unusual in the fact that it contained a stop codon immediately up-stream of the initiation codon for the "mature" core gene. This in-frame ORF is termed the preC region, with the corresponding preC initiation codon at position 3096 (fig 1.3). The other sequenced mammalian Hepadna viruses also contain a preC region in an analogous position to that of HBV. It has been suggested that the core gene of the duck hepatitis B virus is a fusion of the X and core genes to give a single large open reading frame (2893-3021/1-786) (Mandart *et al.*, 1984). This assumption was drawn from the observation that the 5' end of the duck hepatitis virus "core" gene overlapped the 3' end of the polymerase gene, as does the mammalian virus X and polymerase genes. However, Mandart *et al.*, (1984) also report that they could not detect any significant homology between the amino termini of the duck "core" and the mammalian X predicted amino acid sequences.
Examination of the amino acid sequence of the preC regions of the mammalian viruses indicates that it is rich in hydrophobic residues, with the sequence resembling a potential leader signal peptide (Sabatini et al., 1982; Ou et al., 1986). The first 43 amino acids of the duck virus "core" polypeptide is also rich in hydrophobic residues; and with analogy to the mammalian preC region may also represent a leader signal peptide. This would position the start of the "mature" DHBcAg ORF at nucleotide position one on the DHBV sequence. Therefore, the larger size of the duck core gene is possibly due to a larger leader signal peptide coding sequence and probably also a longer carboxy terminal region, since this region also overlaps the polymerase gene to a greater extent than the mammalian core genes.

The nucleocapsid of HBV is composed essentially of a single protein species, the HBcAg protein. The size estimate of the HBcAg protein isolated from viral cores is 19,000–21,000 daltons (Budkowska et al., 1977; Hruska and Robinson, 1977) this is in reasonable agreement with the size prediction of the "mature" core protein. It is not known whether the primary translated core protein product contains the preC region or not. The preC region is not required for expression in mammalian cells (Will et al., 1984; Ou et al., 1986). Ou et al., (1986) have inserted the core coding sequence, with and without the preC region, into an SV40 early gene replacement vector. Transfection of COS cells with the above constructions resulted in the expression of immunologically reactive HBcAg material both with and without the preC region. However, when the preC region was present, the core protein was found to be associated with the endoplasmic reticulum and Golgi structures, indicating that it is membrane associated, possibly due to preC leader signal peptide. The presence of the preC region also resulted in the secretion of HBeAg into the culture medium. In contrast, when the preC region was absent, the core protein was found to be distributed throughout the cytoplasm and no significant HBeAg was secreted from the cells.

The preC region may play a role in the formation of the mature virion. The preC amino acid sequence could anchor the preC/Core protein in the
membranes containing the surface antigen proteins, such that the carboxy terminal portion of the preC/Core faces the cytoplasm. The membrane-bound preC/Core protein could interact with cytoplasmic "immature" core particles, consisting of core proteins lacking the preC region, which have been translated from a mRNA initiating downstream of the preC initiation codon (see section, 1.5.4 The Core mRNAs). This is consistent with the observations of Kamimura et al., (1981) and Yamada et al., (1982), that core particles in HBV infected hepatocytes bud into the cisternae of the endoplasmic reticulum to form Dane particles (Ou et al., 1986).

The HBeAg determinants have been found to be associated with the HBcAg protein. The HBcAg determinant is dependent upon the conformation of the core particles; when the particles are disrupted and the core protein denatured, the antigenic specificity changes from HBcAg to HBeAg (Budkowska et al., 1979; Ohori et al., 1979,1980; Takahashi et al., 1979). Additional evidence that the HBeAg epitope is found on the core protein comes from the work of MacKay et al., (1981), when HBcAg protein produced in E. coli is chemically denatured and partially enzymatically degraded, the HBcAg specificity is converted to immunologically reactive HBeAg. Sequencing of HBeAg protein derived from serum, suggests that the conversion of HBcAg to HBeAg may in part arise by proteolytic cleavage of 33 to 35 amino acids from the carboxy terminus of the HBcAg molecule to give a protein of apparent molecular weight ca. 15,500 daltons (Takahashi et al., 1983).

Pasek et al., (1979) remarked on the protein sequence of the carboxy terminal half of the HBcAg (residues 150-183), due to its protamine like structure with repeating arginine residues separated by serine and proline, suggesting that the C terminal region could bind HBV DNA within the Dane particles. Until recently the possible DNA binding properties of the core protein were unknown. Petit and Pillot, (1985) using a "South-western" technique have demonstrated that the core protein binds 32P-labelled DNA. The DNA binding reaction also seems semi-specific, in that the affinity for binding HBV DNA was stronger than pBR322 DNA.
1.4.3 The Polymerase ORF

The largest open reading frame on the HBV genome is termed the polymerase (pol) gene (407-2902), it has a coding capacity for a polypeptide of 832 amino acids with a molecular weight of ca. 93,500 daltons. The assignment of the endogenous DNA polymerase to this open reading frame was originally based upon purely circumstantial evidence. DNA polymerases of bacteria and viruses have molecular weights in the same size range as the potential product of the pol ORF, therefore, the hypothesis that this open reading frame codes for the endogenous DNA polymerase was proposed (Galibert et al., 1979; Pasek et al., 1979).

The results of Summers and Mason, (1982) suggest that the life cycle of the Hepadna viruses involves replication of the viral DNA via an RNA intermediate using a reverse transcription reaction. This also indicates that the virus probably encodes a reverse transcriptase, most likely the same protein as the endogenous DNA polymerase (see section, 1.6 Viral Replication). Evidence supporting this hypothesis comes from the work of Toh et al., (1983) and Patarca and Haseltine, (1984), in which the predicted amino acid sequence of the pol gene was compared to the predicted amino acid sequences of retroviral and cauliflower mosaic virus reverse transcriptases. The comparisons indicate that the pol gene of HBV contains tracts of amino acid sequence homology with the other viral reverse transcriptases, indicating possible functional relatedness of the viral proteins.

Additional evidence that this ORF encodes a DNA polymerase, is that fusion proteins produced in E. coli containing the pol gene sequences from position 523-1004 fused in frame with E. coli β-galactosidase exhibit an in vitro DNA polymerase activity. This fusion protein and others containing pol gene sequences, have also been found to be reactive by "western" blotting to chimpanzee anti-sera following challenge by HBV, indicating that the pol gene is expressed during HBV infection (McGlynn and Murray, 1987).
A fourth open reading frame is conserved in all mammalian Hepadna viruses, but as yet no functional or structural aspect of the viral life cycle has been associated with the product of this gene. A recent paper of Wollersheim and Hofschneider, (1987) reports that the X gene product may have a trans-activation activity, which suggests a possible involvement of the X protein with the transcription of either viral or cellular genes during the HBV replication cycle. The X gene (2656-3117) has a coding capacity for a protein of 154 amino acids with a molecular weight of ca. 16,500 daltons. The X gene has been cloned and various fusion proteins have been expressed in E. coli (Kay et al., 1985; Moriarty et al., 1985; Elfassi et al., 1986; Meyers et al., 1986; Pugh et al., 1986). The above fusion proteins have been used for immuno-precipitation and "western" blotting analysis to detect antibodies in human anti-sera, specifically reactive to the HBV X gene product, indicating that the X gene product is expressed in natural infections of HBV and constitutes a new HBV serological marker HBxAg.

Recently Chang et al., (1987a) have identified two proteins that reacted by immuno-precipitation with rabbit anti-X-sera, following transfection of HuH-7 cells with cloned HBV DNA. The identified proteins have apparent molecular weights of 17,000 and 22,000 daltons, with the p17 protein corresponding to the predicted size of the X protein. It is not known whether the p22 protein is a HBV encoded derivative of the X protein or a cellular protein co-precipitating with the X protein. Another alternative is that p22 could be a HBV-cellular X fusion protein bearing HBxAg epitope(s). Neither protein was observed in immuno-precipitates of untransfected HuH-7 cell lysates.
1.5 Transcription and Viral Gene Expression

1.5.1 The Viral Transcripts

Only recently has HBV been propagated in cell culture (Sureau et al., 1986). Consequently much of our knowledge of the molecular biology of HBV has been drawn from the analysis of viral gene expression of HBV infected tissues, hepatoma derived cell lines and transfection of non-permissive mammalian cells with cloned HBV DNA.

The HBV genome contained within the Dane particle is a relaxed circular, partially double-stranded molecule, containing a "nick" at a unique site on the long strand. For the viral genome to be transcribed it would be expected that the DNA would have to be "repaired" and made a fully double-stranded covalently closed molecule. An analysis of nuclear DNA from infected tissues, indicates that viral DNA exists as extrachromosomal molecules distinct from the Dane particle DNA. The episomal HBV DNA molecules migrate with apparent molecular sizes of 4.0 kb and 2.2 kb when separated by agarose gel electrophoresis. The two forms of HBV DNA can be converted to a single species with an apparent molecular size of 3.2 kb, upon digestion with Eco RI. With analogy to bacterial plasmids, the 2.2 and 4.0 kb forms represent covalently closed circular supercoiled (cccDNA) and "nicked" open circular (ocDNA) molecules respectively, which are converted to a linear genomic length molecule when digested with Eco RI (Ruiz-Opazo et al., 1982; Elfassi et al., 1984; Miller and Robinson, 1984). The covalently closed circular viral DNA is found predominantly in the nucleus of infected cells, with approximately 50 copies per cell, resulting from de novo DNA synthesis via the reverse transcriptase pathway in the cytoplasm (Tuttleman et al., 1986). The cccDNA is presumed to be the template for transcription of the hepadnavirus mRNAs (Miller and Robinson, 1984; Tuttleman et al., 1986).

Initial studies of HBV gene expression were conducted on a human hepatocellular carcinoma derived cell line PLC/PRF/5 (Alexander cells) (Alexander et al., 1976). The cell line PLC/PRF/5 has been found to
secrete HBsAg 22 nm particles into the culture medium (Marion et al., 1979; Skelly et al., 1979). Analysis of the integrated HBV DNA shows that there are no full length integrated HBV genomes, at least seven HBV fragments, some rearranged, are integrated into the PLC/PRF/5 genome (Brechot et al., 1980; Chakraborty et al., 1980; Edman et al., 1980; Marion et al., 1980b; Shaul et al., 1984; Ziemer et al., 1985). The Alexander cells produce a poly(A)+ mRNA of 2,000–2,500 nucleotides that hybridises to the surface region of the HBV genome (Chakraborty et al., 1980; Edman et al., 1980). This mRNA is approximately the size of the native HBsAg mRNA (see below), however, it has recently been shown that this mRNA is a hybrid molecule containing HBV sequences co-transcribed at its 3′ end with human sequences (Freytag von Loringhoven et al., 1985; Ou and Rutter, 1985) and it is fortuitous that it is approximately the same size as the native HBsAg mRNA.

An approach to the study of HBV gene expression was undertaken. It was reasoned that since the HBV genome is circular; cloned genomic length molecules may not mimic the transcriptional template of HBV. Therefore, multimeric forms of HBV genomes were constructed in vitro, containing complete tandem copies of HBV DNA cloned into E. coli vectors. These constructions could therefore mimic a circular monomeric HBV molecule. Pourcel et al., (1982) using a dimeric construction co-transfected, along with a plasmid containing the Herpes simplex virus type 1 thymidine kinase (tk) gene as a selectable marker, mouse L cells deficient in thymidine kinase (LMtk−). Hypoxanthine [HAT] resistant (tk+) transformants were found to secrete HBsAg into the medium. Three HBV related mRNAs were observed, a major transcript of 2.3 kb which hybridised to the surface region of the HBV genome and two minor transcripts of 4.3 and 0.9 kb, the nature of the minor transcripts was not investigated.

Gough and Murray, (1982) using a similar technique, however, using a construction that contained four "head to tail" copies of HBV (pHBV130.4), co-transformed Rat 2 (tk−) and LMtk− cells. The HAT resistant transformants were found to contain integrated HBV DNA, and to express not only HBsAg but also HBeAg and HBeAg as judged by
radioimmunoassay. The same transformants were later shown to express the HBxAg as assayed by immunofluorescence (Pugh et al., 1986). Therefore, these transformants express all the known HBV related antigens. Gough (1983) investigated the HBV transcripts produced by these cell lines and found four HBV specified mRNAs that hybridise to the coding strand (minus strand); a major 2.4 kb mRNA which hybridises to the surface and X regions of the genome; a 1.0 kb mRNA which hybridises specifically to the X gene and two minor transcripts greater than the genomic length of HBV, of 4.4 and 3.9 kb, which hybridise to all regions of the HBV genome. No transcripts were observed that hybridised to the plus (short) strand or the plasmid vector pBR322. The greater than genomic length mRNAs were found in independently isolated cell lines that express HBCAg and HBeAg and not in cell lines that express only HBsAg. Therefore, the expression of HBCAg/HBeAg probably requires the synthesis of at least one of these greater than genomic length mRNAs. Synthesis of these transcripts would require read-through from one HBV genome to another, therefore mimicking a circular viral genome. The production of these transcripts would also require inefficient polyadenylation, the transcripts only being polyadenylated upon a second (or additional) round of transcription, thus the mRNAs would be terminally redundant.

Cattaneo et al., (1984) isolated RNA from HBV infected chimpanzee liver, and by "northern" analysis of the viral transcripts identified two major transcripts of 2.3 kb and 3.8 ± 0.3 kb, probably corresponding to the 2.4 kb surface mRNA and 3.9 kb mRNAs identified by Gough, (1983). The greater than genomic length mRNA present in infected hepatocytes was found to be much more abundant than in the cell lines constructed by Gough and Murray, (1982). The levels of the 3.8 kb and 2.3 kb mRNAs in infected liver cells were approximately 20 and 50 copies/cell respectively, whereas in the cell lines of Gough and Murray, (1982) the 2.3 kb transcript is in approximately 50 fold excess over the 3.8 kb transcript. This indicates possible differences in the control of viral transcription in primate liver and non-permissive rodent cells. Factors required for optimal transcription of the core mRNA are possibly only present in hepatocytes, allowing optimal expression of the HBCAg/HBeAg.
and viral replication. The 1.0 kb X mRNA was not observed in the preparations of poly(A)% RNA isolated from infected liver.

Recently, using the methods of Pourcel et al., (1982) and Gough and Murray, (1982) as well as transient expression studies, the human hepatitis B virus has been successfully propagated in hepatoma derived cell lines, indicating the requirement of HBV for hepatocyte cells for permissive replication (Sureau et al., 1986; Chang et al., 1987a; Sells et al., 1987; Tsurimoto et al., 1987; Yaginuma et al., 1987b).

1.5.2 The Polyadenylation Signal

Pourcel et al., (1982) and Gough, (1983) both propose that the surface mRNA is polyadenylated at a variant polyadenylation signal at positions 16-21 (TATAAA). The sequence AAUAAA is found 10-30 bases upstream of the site of addition of the poly(A) tail of most eukaryotic mRNAs (Proudfoot and Brownlee, 1976). The sequence at position 16 on the HBV map has only a single mis-match to this sequence and is the preferred candidate for the polyadenylation signal for the HBsAg mRNA. The results of Gough, (1983) indicate that 3' processing of the X mRNA also requires this signal. Cattaneo et al., (1983b), mapped the 3' end of the HBsAg mRNA isolated from infected liver tissue and the cell lines of Gough and Murray, (1982). Using S1 protection studies, a probe from nucleotide 2852 to 430 was protected giving a product of ca. 370 nucleotides, corresponding to a 3' polyadenylation addition site at nucleotide 40 ± 5. This result was repeated in a later study which indicated that the greater than genomic length mRNA is most likely processed at the same polyadenylation signal (Cattaneo et al., 1984) Simonsen and Levinson, (1983) using a simian virus 40 (SV40) based vector containing the HBsAg gene, as well as 5' and 3' flanking sequences (nucleotides 1280 to 84 on the HBV map) positioned behind the SV40 late promoter, expressed transcripts in COS cells and identified species of mRNA directed by the SV40 late promoter. These transcripts were found to terminate within HBV sequences, therefore, utilising HBV
transcriptional termination signals. They isolated cDNA clones corresponding to the HBV containing transcripts and determined the sequence of the 3' ends. The isolated clones contained a poly(A) tract inserted 12 to 19 bases downstream of the sequence TATAAA (nucleotides 16 to 21), with the polyadenylation addition sites between positions 33 to 41, confirming the speculations of Pourcel *et al.*, (1982); Gough, (1983) and Cattaneo, *et al.*, (1983b). The results of Simonsen and Levinson, (1983) also indicate that at least 30 nucleotides downstream of the polyadenylation signal are required for efficient 3' processing.

Recent reports suggest that sequences 3' to the polyadenylation signal of eukaryotic transcripts are necessary for efficient polyadenylation (Hart *et al.*, 1985; McDevitt *et al.*, 1986; Gil and Proudfoot, 1987). One such sequence motif found downstream of the adenovirus E2A polyadenylation site, TT(A/G)TTTTT (U-rich element) (Hart *et al.*, 1985; McDevitt *et al.*, 1986) shares nucleotide homology to a sequence 28 bases downstream of the HBV polyadenylation signal (TcGTTTTT, position 49 to 56) (unpublished observations). Evidence that this sequence could be involved in the polyadenylation of the HBV transcripts comes from the above results of Simonsen and Levinson, (1983). Three prime deletion mutants, with different lengths of sequence downstream of the polyadenylation signal were tested for their efficiency of polyadenylation. One such deletion derivative, pHBVterm 13, deleted for sequences downstream to position 52 contains the polyadenylation signal with the deletion point within the U-rich sequence defined above. Transcripts directed by this construction in COS cells were inefficiently polyadenylated. A similar construction, pHBVterm 14, deleted for sequences downstream to position 67, contains both the polyadenylation signal and the U-rich element. This deletion derivative directed the synthesis of mRNA of the same size and efficiency as their standard construction (pDL RI), which contains HBV sequences from the *Eco* RI site (position 1280) downstream to the *Bgl* II site (position 84). This strongly suggests that this U-rich element could be involved in 3' processing of the HBV transcripts. Related sequences are found downstream of the polyadenylation signals of the other Hepadnaviruses,
indicating possible conservation of a functionally important DNA/RNA sequence.

The single mis-matches between the HBV polyadenylation signal, the U-rich element and the consensus sequences, may permit read through transcription of the polyadenylation signals, due to a relatively less specific interaction with the polyadenylation machinery. Therefore allowing the synthesis of the greater than genomic length mRNAs observed by Gough, (1983) and Cattaneo et al., (1984). Saito et al., (1986) using an Adenovirus type 5 based vector, containing an HBV Bgl II fragment (523-84), transfected HeLa cells and expressed HBV transcripts. Two-thirds of the viral transcripts failed to terminate at the HBV polyadenylation signal and proceeded into vector sequences indicating that the HBV polyadenylation signals are functionally "leaky".

1.5.3 The Surface mRNAs

The HBV surface mRNA has an apparent molecular size of ca. 2,300 nucleotides as determined by "northern" hybridisation (Pourcel et al., 1982; Gough, 1983; Cattaneo et al., 1984). Both Pourcel et al., (1982) and Gough, (1983), by inference to the length of the viral transcripts and on the basis of hybridisation studies postulated, that the surface antigen promoter is contained upstream to the preS region. A "TATA"-like sequence at position 876 (TATATAAA) has been proposed to be the promoter, implying the start of the transcript at around 900-910 (Pourcel et al., 1982). Gough, (1983), however, observed hybridisation of the transcript to a probe extending from 523-900, and therefore proposed that the surface promoter may be at position 790 (TATATAAA) or 845 (TATTTA), with initiation sites 815-825 or 870-880 respectively.

The original studies conducted to investigate the transcriptional promoter signals of the preS/surface gene were those of Rall et al., (1983). These studies involved the generation of in vitro run off transcripts using HeLa whole cell extracts. Rall et al., (1983) identified an RNA polymerase II dependent surface transcript initiating
at position 907 ± 10 on the HBV map, 25 nucleotides downstream of a potential "TATA"-like sequence TATATAA (876-882), in agreement with the site postulated by Pourcel et al., (1982).

Cattaneo et al., (1983b) isolated surface mRNA from HBV infected chimpanzee liver and the Rat2/130.4/TK4 cell line of Gough and Murray, (1982). Using an S1 protection procedure, with a probe covering sequences 533-1409, the 5' end of the HBsAg mRNA was mapped to position 1255 on the HBV map, 185 nucleotides upstream of the S gene coding sequence. The same site was mapped when the RNA/DNA hybrids were digested with Exo VII or analysed by primer extension experiments, indicating that this site does not represent a splice site and no large intron exists in the 5' end of the surface transcript. This result was unexpected considering the in vitro transcription results of Rall et al., (1983) and the hybridisation results of Gough, (1983) which map the 5' end of the surface mRNA upstream of the preS region. The 5' initiation site mapped by Cattaneo et al., (1983b) is 17 nucleotides upstream from the initiation codon for the preS2 open reading frame and could therefore code for both the preS2 and surface proteins.

No TATA box promoter is found in the expected position 20-30 nucleotides upstream to position 1255. A sequence found 30 nucleotides upstream of the SV40 late mRNA initiation site has been defined as being important in regulation of transcription of the SV40 late mRNA (Brady et al., 1982). Cattaneo et al., (1983b) observed that a sequence upstream of the surface mRNA initiation site shares extensive homology to the sequence defined by Brady et al., (1982). A second sequence motif with homology to the SV40 origin of replication, which has also been shown to be involved in the regulation of SV40 late transcription (Contreras et al., 1982; Cattaneo et al., 1983a), is found 60 nucleotides upstream of the surface mRNA start site (Cattaneo et al., 1983b). It should be noted that an additional consensus sequence for the binding site of the transcription factor CTF (CCAAT), which is a common element of many eukaryotic promoters (Benoist et al., 1980; Efstratiadis et al., 1980), is found between the homology to the SV40 origin and late promoter.
regions, 51 nucleotides upstream of the surface mRNA start site (unpublished observation) (fig 1.7).

Since the above studies were reported, the expression of HBsAg gene and its mRNA have been studied in a variety of expression systems. These expression systems include the study of stably transformed rodent cells and hepatoma cell lines containing integrated HBV DNA, HBV infected human and chimpanzee liver tissue and transient expression systems using viral based vectors containing the HBsAg gene. The results of some of these experiments are presented in table 1.3. Summation of these results indicates that the major surface mRNAs exist as a closely related set of RNA species with heterogeneous 5' ends. Probably only three major species of surface mRNA exist with initiation sites between positions 1250-1260, 1274-1280 and 1284-1290. The most 5' initiation site maps upstream of the preS2 initiation codon with the two smaller mRNA initiating 3' to the preS2 ATG codon, 145-165 nucleotides upstream of the surface ATG initiation codon. The transcriptional initiation site, position 907, observed in vitro (Rall et al., 1983), is utilised in mouse fibroblasts, infected liver and in hepatoma derived cell lines, however, at a very low efficiency. The relative abundance of this preS1 mRNA produced in the cell line PLC/PRF/5 has been estimated at only 2% of the total surface mRNA (Ou and Rutter, 1985). Another low abundance preS1 mRNA is expressed in the cell line L/130.4/TK154, constructed by Gough and Murray, (1982); Cattaneo et al., (1984) observed the mRNA initiating at position 851±20, in the mouse L cell derived cell line. The existence of this mRNA would account for the hybridisation results of Gough (1983) in which surface mRNA was found to hybridise to probes extending from 523 to 900.

The surface mRNAs are probably directed by two separate promoters, the preS1 "TATA" box promoter (SPI) at position 876 to 882 and the preS2/surface "SV40-late-like" (SPII) promoter situated between positions 1188 and 1260. The downstream initiation sites could be derived by some form of 5' processing, however, no introns are present in the 5' end of the major transcripts (Cattaneo et al., 1983b; Standring et al., 1984).
Figure 1.7 Alignment of nucleotides from the HBV major surface promoter (HBV130 sequence; Pugh et al., 1986) (middle), the SV40 late promoter region (bottom) and the SV40 origin (top). The sequence important in positioning the SV40 major late transcript (Brady et al., 1982) is underlined. The 4-bp deletion in the SV40 origin strongly reducing SV40 late transcription (Contreras et al., 1982) is overlined. A putative CTF binding site (CCAAT) is underlined on the HBV sequence. The major initiation of transcription sites are indicated with a *, homologous nucleotides between the SV40 and HBV sequences are indicated with a +. Figure adapted from Cattaneo et al., (1983b).
<table>
<thead>
<tr>
<th>Cell type</th>
<th>Mapping method</th>
<th>Transcription initiation sites</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>In vitro COS</td>
<td>Run off trans.</td>
<td>907±10</td>
<td>Rall et al., (1983)</td>
</tr>
<tr>
<td>Chimpanzee liver</td>
<td>SI, ExoVII, PE</td>
<td>1255</td>
<td>Cattaneo et al., (1983b)</td>
</tr>
<tr>
<td>PLC/PRF/5</td>
<td>PE</td>
<td>907±4 (2%)</td>
<td>Ou and Rutter, (1985)</td>
</tr>
<tr>
<td>HeLa</td>
<td>SI</td>
<td>908</td>
<td>Saito et al., (1986)</td>
</tr>
<tr>
<td>Rat fibroblast</td>
<td>SI, PE</td>
<td>885-890</td>
<td>Siddiqui et al., (1986)</td>
</tr>
<tr>
<td>HuH-7</td>
<td>SI</td>
<td>907±4</td>
<td>Yaginuma et al., (1987b)</td>
</tr>
<tr>
<td>Mouse 3T3</td>
<td>SI, PE</td>
<td>1251</td>
<td>Zelent et al., (1987)</td>
</tr>
<tr>
<td>Chimpanzee Liver</td>
<td>SI, PE</td>
<td>910±20</td>
<td>Will et al., (1987)</td>
</tr>
</tbody>
</table>

§ Run off trans. = *in vitro* run off transcription; SI, ExoVII = SI, ExoVII digestion mapping, PE = primer extension mapping
The identity of the preS2/surface promoter has been confirmed by cloning fragments of HBV DNA devoid of the SPI promoter and expressing HBsAg in mammalian cells under the control of the SPII promoter of HBV (Standring et al., 1984; Siddiqui et al., 1986). Siddiqui et al., (1986) using a heterologous system with the surface promoter (SPII) directing the expression of bacterial chloramphenicol acetyltransferase (CAT) in CV-1 cells, defined the major HBsAg promoter as being between nucleotides 993 and 1310 on the HBV map. A recent report of Shaul et al., (1986a) using a heterologous system similar to Siddiqui et al., (1986), indicates that the proximal promoter elements defined by Cattaneo et al., (1983b) are possibly dependent on at least one upstream distal element for maximal expression. Shaul et al., (1986a) identified a binding site for nuclear factor 1 (NF-1), TGG N GCCAA, at positions 1068 to 1081. The expression of 5' deletion mutants in HeLa cells indicate that deletions extending to position 1069 on the HBV map resulted in a 5-fold reduction in CAT expression. Therefore, Shaul et al., (1986a) propose that NF-1 is a transcriptional regulatory factor for the major surface promoter.

Three separate proteins are translated from the preS/surface ORF. It was originally proposed that the entire open reading frame was translated to give a precursor polypeptide which was subsequently processed to the major surface proteins (p25/gp28) and the preS proteins. However, this idea was refuted when Persing et al., (1985) introduced a +1 frame shift mutation at the Eco R1 site (1280) between the start of the coding sequences for the preS2 and surface proteins. Mouse L cells transfected with this construction secreted HBsAg 22 nm particles. Analysis of the protein composition indicated that the 22 nm particles were devoid of the preS derived proteins and contained only the smaller major surface proteins. These results indicate that the surface protein is translated independently of the preS proteins and is not derived from cleavage of a precursor protein. A more plausible explanation for the translation of the HBsAg polypeptides is that each protein is translated from separate species of mRNA. The preS1 protein is translated from an mRNA, initiating at position 907, which is directed by the SPI promoter. While the preS2 and major surface proteins are
translated from mRNAs directed by the SPII promoter. The preS2 protein is probably translated from a mRNA which initiates between positions 1250-1260, while the major surface protein is translated from two alternative mRNAs initiating between positions 1274-1280 and 1284-1290 (fig 1.8).

15.4 The Core mRNAs

The core mRNA (greater than genomic length mRNA) has not been as well characterised as the surface mRNA due to its low abundance in most expression systems. The core mRNA isolated from infected liver has an apparent size of ca. 3,800 ± 300 nucleotides (Cattaneo et al., 1984). The major 5' end of liver derived core mRNA has been mapped to position 3100 ± 2 on the HBV map; minor transcripts initiating upstream to position 3100 were also observed (Will et al., 1987). The same initiation site was mapped by Saito et al., (1986) using an Adenovirus vector system although the 5' end appeared to be heterogeneous. Zelent et al., (1987) using a head to tail dimer of HBV, transfected mouse 3T3 cells and expressed HBV transcripts. The major core transcripts appeared to have 5' ends at positions 3027 ± 5 and 2902 ± 5 with minor mRNAs initiating at approximate positions 3040, 3060, 3070 and 3097. Yaginuma et al., (1987b) have used a transient assay system to produce Dane particles in hepatocellular carcinoma derived cell lines. The core mRNAs produced in these studies had 5' initiation sites at positions 3063/4, 3072 ± 1 and 3099 ± 1 in agreement with both the major core mRNA of Will et al., (1987) and the minor transcripts of Zelent et al., (1987) (fig 1.8). The above difference in the map positions of the 5' ends is possibly due to different transcriptional affinities of hepatocyte derived and non-permissive mammalian cells, as reflected by the greater abundance of the core mRNA in infected liver tissue (Cattaneo et al., 1984). The 5' ends of the core mRNAs of DHBV, GSHV and WHV have been mapped to approximately the same positions as the HBV core mRNAs on their respective maps (Buscher et al., 1985; Enders et al.,1985; Moroy et al., 1985). No promoter sequences have been mapped immediately upstream of the initiation sites of the HBV core mRNAs and
Figure 1.8 Diagram of the HBV transcriptional regulatory region. The transcriptional regulatory signals (top) are shown in relation to the start of the viral open reading frames. Open arrows indicate the transcriptional start sites, positions of the transcriptional start sites are taken from Yaginuma et al., (1987b) and Treinin and Laub, (1987). The positions of the protein binding sites (NF-1, U1-U3, E) are indicated above the line. The start and stop codon positions for the open reading frames are shown below.
no "TATA" consensus is found 20 to 30 bases upstream of the mapped 5' ends, and like the major surface promoter the core promoter probably represents non-TATA promoter.

The major 5' end of the core mRNA (position 3100 ± 2) mapped by Will et al., (1987), initiates 5 nucleotides downstream of the preC ATG initiation codon, therefore, the mRNA could only code for the "mature" core protein. The preC region could be translated from the mRNA species initiating upstream to the preC ATG codon at positions 3063/4, 3072 ± 1 (Yaginuma et al., 1987b) (fig 1.8). Such mRNAs have also been observed in GSHV, and WHV infected liver (Enders et al., 1985; Moroy et al., 1985).

No HBV mRNA has been observed to initiate immediately upstream of the polymerase gene; it is proposed that the core mRNA also encodes for this protein. Translation of the polymerase protein would require that the mRNA be polycistronic, with the polymerase protein being synthesised by internal initiation of translation (reviewed by Kozak, 1986). Alternatively the core mRNA could be spliced to give minor transcripts in which the pol initiation codon is the first available AUG on the mRNA, thus allowing synthesis of the polymerase protein. However, no spliced HBV transcripts have been observed, although this may be due to the low abundance of these mRNAs. An alternative hypothesis has been proposed by Will et al., (1986). The arrangement of the HBV core and polymerase genes is similar to the arrangement of the analogous retroviral gag and pol genes. Therefore, the HBV polymerase/reverse transcriptase could be synthesised via a HBcAg-Pol (c-pol) fusion protein reminiscent of the retroviral gag-pol polyprotein produced by ribosomal frameshifting during translation (Jacks and Varmus, 1985). However, it should be noted that the HBV pol gene is in the +1 frame relative to the core gene, whereas most retroviral pol genes are in the -1 frame relative to the gag gene.

In support of their hypothesis Will et al., (1986) have observed c-pol fusion proteins in liver samples from patients with primary liver cancer. The tumours contained integrated HBV DNA, with no evidence of active HBV replication. It is possible that these fusion proteins could result from
rearrangement of the integrated HBV DNA, with the resultant transcription/translation of fused core and polymerase genes. Will et al., (1986) argue against this possibility, since they observed that separate isolates of the tumours expressed a 35,000 dalton fusion protein, indicating that the separate tumours synthesise either the same or similar protein which would not be expected if the fusions were due to aberrant re-arrangement of the integrated HBV DNA. Will et al., (1986) suggest that the sizes of the observed c-pol fusion proteins could result from incomplete synthesis or processing of the HBV reverse transcriptase in the non-HBV-replicating cells. It remains to be conclusively shown whether these c-pol fusions are produced during natural HBV infection. Detection of the native polymerase protein should answer this question.

Miller, (1987) has observed that the predicted amino acid sequence of the core protein contains a stretch of amino acids with homology to protease active site sequences of cellular and viral proteases. Miller, (1987) proposes that the HBV putative protease-like sequence (amino acid residues 27-37 of the "mature" core protein) could act by proteolytic self-cleavage of the core protein to generate the serum HBeAg. Two possible pathways could determine the fate of the core protein. In the first pathway the protease is in an inactive form and no cleavage takes place. After translation the core protein remains intact and is incorporated into the nucleocapsid of the virus. In the second pathway the protease becomes activated (possibly by post-translational modification), and the core protein is cleaved to remove the carboxy terminal 33-35 amino acids (Takahashi et al., 1983). This reaction would generate HBeAg (from the amino terminus) and a small protein containing the protamine-like region (from the carboxy terminus) of the core protein. The HBeAg is then secreted from the infected cell where it is found as a free protein in the serum. Miller (1987) also proposes that the protease activity of the HBV core protein may be necessary for post-translational modification of the putative c-pol polyprotein to form an active polymerase similar to the proteolytic cleavage of the retroviral gag-pol polyprotein. Therefore, the generation of the HBeAg protein could possibly be a by-product of the HBV replication system. If this is
the case, and depending upon where the translational ribosomal frame shifting occurs, it is possible that an active polymerase derived from the putative c-pol fusion protein could contain the protamine-like amino acid sequence at its amino terminus and therefore may have nucleic acid binding properties.

1.5.5 The HBV Enhancer Element

Sequences have been found on the HBV genome which can effect the expression of the core mRNA. Shaul et al., (1985) inserted a fragment of HBV (2243-2964) upstream of the CAT gene in a SV40 based vector. Significant CAT expression was observed in transfected Alexander cells when the core promoter region was inserted in the sense orientation to the CAT gene. When a sub-fragment (2682-2964) of this region was inserted in sense to the CAT gene, no expression was observed which indicates that sequences between 2243 and 2682 influence the expression of the core promoter. Shaul et al., (1985) propose that the sequences 2243 to 2682 contain an enhancer element. When the enhancer fragment was inverted with respect to the core promoter fragment, significant CAT expression was also observed, indicating that the enhancer element is independent of orientation. The HBV enhancer was also tested in a heterologous system; when an HBV fragment (2015-2519) was inserted upstream of the SV40 early promoter and used to direct CAT gene expression in COS cells, the CAT expression was induced 105 fold above the expression level of the vector containing no enhancer. Using a series of deletions, the HBV enhancer activity (although at a reduced efficiency) was found to reside between nucleotides 2360 to 2519. Insertion of the enhancer region 3' to the CAT gene also resulted in increased expression from the SV40 early promoter. The activity of the HBV enhancer also appears to be semi-tissue specific, in that the effect of the HBV enhancer upon the HBV core promoter expression of the CAT gene was 50-fold higher in PLC/PRF/5 cells as compared to expression in CV-1 cells. This tissue specific enhancer activity may explain the low expression level of the core mRNA in non-permissive mammalian cells and
the relatively high expression levels in HBV infected liver (Gough, 1983; Cattaneo et al., 1984).

It should be noted that the core promoter DNA fragments used by Shaul et al., (1985) only contain sequences upstream of position 2964; the mapped core mRNA initiation sites are positioned 99-136 bases downstream of this site. It is possible that the constructions used by Shaul et al., (1985) may not contain the complete core promoter sequences, although they do contain the RNA polymerase II "TATA" promoter (position 2933-3939) found to be weakly active in vitro, with the probable transcriptional initiation site at 2961/2 (Rall et al., 1983). However this promoter and mRNA initiation site are probably not utilised in vivo (Will et al., 1987; Yaginuma et al., 1987b).

Tognoni et al., (1985) have also identified the enhancer element between nucleotides 2400-2600. The exact nucleotide sequence of the HBV enhancer element is not known. Tognoni et al., (1985) could not detect any perfect matches to consensus sequences of other enhancer elements, although they do report a run of alternating purines and pyrimidines between positions 2513 and 2523, and suggest that these sequences could be involved in enhancer activity.

Shaul and Ben-Levy, (1987) using protein-DNA binding filter assays, have identified DNA protein binding sites between nucleotides 2107 and 2655 on the HBV map. The protein binding sites were further analysed by DNase I footprinting studies. Multiple binding sites were observed at positions 2161-2180 (UE3); 2252-2271 (UE2); 2309-2328 (UE1) and 2460-2486 (E). The protein binding sites UE1, UE2 and UE3 map upstream of the defined enhancer element (2400-2600). Factors that bind to these three sites were only found in nuclear extracts of fully differentiated hepatocyte derived cells and not in non-hepatocyte cells (eg. Wish cells or HeLa cells). Evidence that these binding sites could be involved in enhancer expression comes from the results of Shaul et al., (1985); deletion of the sequences from 2015-2360 resulted in a reduction of the enhancer activity, suggesting that the factors that bind to the upstream sites could act in a synergistic manner with the defined enhancer.
region. If these binding proteins act in such a way, this may explain the semi-tissue specificity of the HBV enhancer element. The E binding site maps within the enhancer region. The E binding factor was found not only in hepatocyte derived cells but also in HeLa cells. Shaul and Ben-Levy, (1987) observed a DNase I hypersensitive site at the 5' end of the E region, when using nuclear extracts from hepatocyte derived cells in their DNase I footprinting studies. When using HeLa cell extracts the E region was protected indicating the presence of the E binding factor, however, the DNase I hypersensitive site was not observed. Shaul and Ben-Levy, (1987) suggest that the E factor of HeLa cells may be related, but not identical to that found in liver cells. It remains to be determined whether these protein binding sequences are involved in enhancer function. Mutation and expression studies need to be conducted in order to confirm this.

The HBV enhancer region is situated about 1150 nucleotides downstream of the major initiation site for the surface mRNA; and 500 nucleotides upstream of the major initiation sites for the core mRNA. Since enhancer elements are known to have an effect when positioned downstream of a gene, and the HBV enhancer fulfils this requirement (Shaul et al., 1985), it is conceivable that the enhancer may also augment expression from the surface promoter. Indications for this come from the studies of Chang et al., (1987b). They used a heterologous expression system with the surface promoter region (523-1280) upstream of the CAT gene and the HBV enhancer region (2243-2682) inserted upstream to the surface promoter. Chang et al., (1987b) demonstrated that CAT expression was increased 20 fold in the hepatoma cell line HuH-7, as compared to expression using constructions lacking the HBV enhancer but still retaining the surface promoter. Although the system of Chang et al., (1987b) was artificial in respect to the native positions of the surface promoter and the enhancer, their results demonstrate that the enhancer can interact with the surface promoter and augment expression. However, it remains to be determined whether the preS1 or the major surface promoter, or both, are stimulated by the effect of the enhancer when the elements are in their native conformation.
A glucocorticoid-responsive element has also been identified on the HBV genome. Tur-Kaspa et al., (1986) have shown, using a heterologous expression system containing the SV40 early promoter, the CAT gene and fragments of HBV containing the HBV enhancer and upstream sequences, that expression of the CAT gene in hepatoma derived and non hepatic derived cell lines can be enhanced above the level due to the effect of the enhancer by the addition of dexamethasone (Dex) to the culture medium. Dex augmentation of CAT expression, is due to a glucocorticoid-responsive element which was shown to be distinct from the previously mapped enhancer. The glucocorticoid-responsive element was found to reside between nucleotides 1310 and 2016, possibly the sequence CCAACTTGTCCT (1635-1646) which has homology to the consensus sequence GNNACAANNTGTYCT found in the human growth hormone gene which functions as a specific glucocorticoid receptor binding site (Moore et al., 1985; Tur-Kaspa et al., 1986). It is not known whether the hormone-responsive element of the HBV genome has enhancer activity independent of the HBV enhancer defined above, or, whether the effect is mediated via the known enhancer (Tur-Kaspa et al., 1986).

1.5.6 The X mRNA

An X specific mRNA has been observed to be expressed in non-permissive mammalian cells transformed with HBV DNA. The size estimates of this message range from 700 to 1,000 nucleotides (Gough, 1983; Saito, et al., 1986; Zelent et al., 1987). The 5' ends of this mRNA were determined by Saito et al., (1986), at least 10 discrete 5' ends were observed initiating between positions 2530 and 2580. This result is unexpected, considering the results of Gough, (1983) who observed that a probe from 2271-2512 hybridised to the X mRNA. On the basis of these hybridisation studies and size estimates of the X mRNA, Gough, (1983) proposed that the X mRNA promoter could be around position 2300. It is possible that Gough, (1983) was detecting the presence of minor X transcripts as with the case of the preS/surface mRNA, or that the different expression systems direct the synthesis of different X mRNAs.
The HBV enhancer region is positioned 5' to the X gene and could therefore also control the expression of the X mRNA. Evidence for this comes from a recent report of Treinin and Laub, (1987); using a heterologous system with the HBV sequences 2244-2635, bearing the enhancer element and sequences upstream of the X gene, inserted in the sense orientation upstream of the CAT gene, observed CAT expression in transfected HepSK cells. The 5' ends of the X promoter directed transcripts were mapped by primer extension experiments. The 5' ends appear to be heterologous with three major transcripts and possibly four minor transcripts having multiple start sites spanning nucleotides 2530 to 2630, in agreement with Saito et al., (1986). The three major initiation sites were mapped to nucleotide positions 2590 ± 5, 2613 ± 5 and 2620 ± 5 on the HBV map (fig 1.8). Deletion of HBV sequences 2250-2396 resulted in a 12-fold reduction of CAT expression in Alexander cells; deletion of part of the enhancer region, from nucleotides 2244-2497 abolished CAT expression (Treinin and Laub, 1987). The above experiments have been repeated using similar constructions expressed in HepG2 cells (Siddiqui et al., 1987). The major X mRNA 5' ends were mapped to positions 2535, 2560,2580 and 2590, a slight difference to the 5' ends mapped by Treinin and Laub, (1987) possibly reflecting the large number of heterogeneous initiation sites directed by this promoter.

Deletion of sequences upstream to the defined enhancer region resulted in a decrease of X promoter directed CAT expression (see above). The results of Shaul et al., (1985) indicate that the maximal enhancer effect is also observed when sequences upstream of position 2360 are present. Siddiqui et al., (1987), have shown that expression of the X promoter in HeLa cells was less than 10% as efficient as in HepG2 cells, indicating a tissue-specific character. The above results indicate that the X gene promoter requires the activity of the enhancer for its function, however the possibility that the X promoter and the enhancer are the same element, or two independent overlapping elements, remains to be determined.

The X mRNA has only been observed by expression of HBV DNA in non-permissive mammalian cells; it has not been observed in HBV infected
liver (Cattaneo et al., 1984; Will et al., 1987) nor in GSHV or WHV infected liver tissues (Enders et al., 1985; Moroy et al., 1985). It is possible that expression of this message is an artifact of the expression systems used, however, the results of Treinin and Laub, (1987) and Siddiqui et al., (1987) indicate that the X promoter/enhancer is active in hepatoma derived cell lines. Therefore, the X mRNA has the potential to be expressed in HBV infected hepatocytes. Failure to detect the X mRNA in infected tissues could possibly be due to the X mRNA being a low abundance message or that it is transiently expressed during HBV infection. Serological evidence indicates that the X protein is expressed during infection (Kay et al., 1985; Moriarty et al., 1985; Elfassi et al., 1986; Meyers et al., 1986), supporting the possibility of the X mRNA being transcribed, assuming that the S mRNA or other HBV mRNAs do not function as polycistronic messages with regard to the translation of the X protein.

1.6 Viral Replication

The direct study of HBV DNA replication has been limited due to, until recently, the inability to propagate the virus in tissue culture and the limited availability of infected tissue samples. Much of our knowledge of hepadnaviral replication has been drawn from the studies of the animal virus models. The results of these studies have been extrapolated to include HBV by analysis of the structural features of the HBV transcripts and virion DNA.

The first insight into the replication mechanisms of the hepatitis B viruses came from the work of Summers and Mason, (1982). Isolated "immature" core particles from DHBV infected liver were shown to have two DNA polymerase activities; in contrast to the mature virion particles both plus and minus strands appeared to be labelled in vitro, in the presence of $^{32}$P-dNTPs. The synthesis of the viral DNA strands was found to be asymmetric; the minus strand being synthesised via a reverse transcription activity utilising an RNA template, with the RNA being degraded by an RNAase H activity as it is reverse transcribed into DNA.
Electrophoresis and "Southern" blot analysis of total DNA from Hepadnavirus infected cells has shown the presence of a heterogeneous population of rapidly migrating species of DNA. Hybridisation with strand-specific probes has shown that the heterogeneous species are predominantly single-stranded minus strands, shorter than, and equal to, unit length viral DNA. These species represent intermediates and end products of the reverse transcription reaction, with little or no associated plus strand DNA (Mason et al., 1982; Molnar-Kimber et al., 1983; Weiser et al., 1983; Blum et al., 1984). Summers and Mason (1982) also demonstrated that plus strand synthesis utilises the viral minus strand as a template, with the plus strand being fully annealed to its minus strand complement.

The most likely candidate for the RNA pregenome is the greater than genomic length mRNA observed in poly(A)^+ fractions of infected liver tissue (Cattaneo et al., 1984; Buscher et al., 1985; Enders et al., 1985; Moroy et al., 1985). The 5' ends of the pregenome/core mRNA appear to be heterogeneous in GSHV and WHV infected liver (Enders et al., 1985; Moroy et al., 1985) and possibly also in HBV infected cells (Will et al., 1987; Yaginuma, et al., 1987b). Enders et al., (1987) have analysed GSHV RNAs in cytoplasmic lysates, they observed that the 2.3 kb surface mRNA and the two largest core mRNAs (initiating at positions 3205 ± 4 and 3198 ± 4) appear to be associated with polyribosomes, indicating that they were being translated to yield the surface and the preC/core proteins. In contrast most of the shortest pregenomic mRNA (initiating at position 3227 ± 3) was found to be primarily associated with viral core particles. The ratio of this RNA to other viral RNAs was 100 to 1 within the core particle, therefore, this shortest pregenomic mRNA is preferential packaged within the particles. The 5' end of the packaged pregenomic RNA of GSHV corresponds to the major 5' ends of the pregenomic mRNAs observed in HBV, WHV and DHBV infected liver (Will et al., 1987; Moroy et al., 1985; Buscher et al., 1985).

The smallest detectable nascent minus DNA strand (ca. 30 bases) has been shown to have a covalently attached protein at its 5' end, which possibly acts as a primer for minus strand synthesis (Molnar-Kimber et al., 1983).
Priming of the Hepadnavirus minus strand synthesis has been shown to occur within a sequence termed DR1 (Direct Repeat 1) (fig. 1.9) present within the terminal redundancy [R], found on both the 5' and 3' ends of the pregenomic RNA (HBV sequences 3100 ± 2 to 40 ± 5) (Molnar-Kimber et al., 1983; Seeger et al., 1986; Will et al., 1987). The 5' end of the HBV minus strand has been mapped to position 3108 (Will et al., 1987), three nucleotides from the 5' boundary of the DR1 sequence (3106-3116).

Termination of the minus strand synthesis could occur in response to some internal signal, or by exhaustion of the pregenomic template. The 3' end of the GSHV minus strand has been mapped to position 3227, corresponding to the 5' end of the shortest major pregenomic RNA (Seeger et al., 1986). This result, if taken in conjunction with those of Enders et al., (1987), indicates that the reverse transcription reaction most likely proceeds to the end of the pregenomic RNA. Placement of the GSHV minus strand 5' end at 3235 and the 3' end at position 3227, would mean that the minus strand of the virion DNA is greater than genomic length with a short terminal redundancy [r] of 9 nucleotides. By analogy to the GSHV the minus strand of HBV would also have a terminal redundancy [r] of 9 nucleotides (positions 3100-3108) (Will et al., 1987).

The 5' end of the virion plus strand appears to be homogeneous and probably reflects the priming site for the second strand synthesis. An oligoribonucleotide (17-19 mer) has been found to be attached to the 5' end of the Hepadnavirus plus strands. This RNA molecule most likely serves as a primer for plus strand synthesis, with the plus strand DNA sequence beginning at the 3' boundary of the viral DR2 sequence (Direct repeat 2) (fig. 1.9) (Lien et al., 1986; Seeger et al., 1986; Will et al., 1987). The HBV plus strand contains a RNA primer of ca. 17 bases, with the plus strand DNA sequence beginning at position 2882 ± 1, at the 3' end of the DR2 sequence (2872-2882) on the HBV map (Will et al., 1987). Lien et al., (1986) and Seeger et al., (1986) have investigated the origin of the RNA primer; their RNA sequencing results indicate that the oligoribonucleotide is derived from the DR1 sequence, which is transposed to the DR2 position. The RNA primer could be derived from either DR1 sequence represented within the 5' and 3' [R] regions of the pregenome.
a) DR1 Sequences

\[
\begin{align*}
\text{DR1} & \\
\text{AACTTTTTCACCTGTGC} & \text{HBV (3106-3116)} \\
\text{ATCTTTTTCACCTTGC} & \text{GSHV (3232-3242)} \\
\text{ATCTTTTTCACCTGTG} & \text{WHV (3230-3239)} \\
\text{AGAATTACACCCCTCTC} & \text{DHBV (2910-2921)}
\end{align*}
\]

b) DR2 Sequence

\[
\begin{align*}
\text{DR2} & \\
\text{TTCACCTCTGC} & \text{HBV (2872-2882)} \\
\text{TTCACCTGTCG} & \text{GSHV (3010-3020)} \\
\text{TCACCTGTCG} & \text{WHV (3008-3017)} \\
\text{TACACCCCTCTC} & \text{DHBV (2852-2873)}
\end{align*}
\]

Figure 1.9 Nucleotide sequence (plus strand) of the hepadnavirus direct repeats (DR1 and DR2) showing their respective nucleotide positions.

a) The DR1 sequence (underlined) showing the upstream terminal redundancy (r) found on the 5' and 3' end of the minus strand DNA.

b) The DR2 sequence.
RNA, however, evidence suggests that the RNA primer is derived from the 5' end of the pregenomic RNA. Lien et al., (1986) observed that the DHBV RNA primer is capped, and that the 5' end of the packaged pregenomic RNA corresponds to the 5' end of the RNA primer (Lien et al., 1986; Seeger et al., 1986; Enders et al., 1987).

A model for Hepadnavirus replication (fig. 1.10) has been proposed based on the above experiments (Summers and Mason, 1982; Molnar-Kimber et al., 1983; Buscher et al., 1985; Lien et al., 1986; Seeger et al., 1986; Will et al., 1987). For simplicity HBV is used as an example here. Upon entry to a susceptible cell the virion DNA is made a fully double stranded covalently closed circle and transported to the nucleus of the cell where it is transcribed to give the viral mRNAs. The mRNAs are translated in the cytoplasm to generate the core, polymerase and possibly the surface and X proteins. In the cytoplasm the pregenome RNA is packaged within immature core particle along with the HBV polymerase protein (reverse transcriptase). Priming of the minus strand occurs in the terminal redundancy [R] within the sequence DR1, presumably by the uncharacterised protein attached to the 5' end of the minus strand. Initiation of the minus strand could occur at either the 5' or 3' [R] of the RNA template; if priming occurs at the 5' [R], elongation of the minus strand would require a template switch to the 3' end of the pregenomic RNA. Therefore, for simplicity, priming at the 3' [R] sequence is shown in fig. 1.10. Initiation of reverse transcription occurs at position 3108 and proceeds to the end of the RNA template terminating at position 3100, generating a short terminal redundancy [r] of 9 nucleotides at the 5' and 3' ends of the minus DNA strand. The RNA template is probably degraded by an RNAase H activity behind the growing point of the nascent minus strand.

The second strand synthesis is primed by an approximately 17-base RNA oligomer containing the DR1 sequence and flanking upstream sequences extending to position 3100 ± 2 which is released from the 5' end of the viral RNA and transported by undefined mechanisms to the position of the DR2 sequence on the newly synthesised minus strand. Synthesis of the plus strand is initiated at position 2882 ± 1 and proceeds to the 5' end
of the minus strand containing the \([r]\) sequence, with the template being exhausted shortly after initiation resulting in a strong stop plus phenomenon as observed in retrovirus replication (Smith et al., 1984; Champoux et al., 1984). Evidence for a strong stop plus phenomenon for Hepadnaviruses comes from the results of Molnar-Kimber et al., (1983), who have shown the existence of an actinomycin D sensitive DNA product of ca. 68 nucleotides present in DHBV liver derived core particles. The size of this species of DNA corresponds to the size of the DHBV cohesive overlap, strongly suggesting that it represents a strong stop plus product of DHBV. Further extension of plus strand synthesis would require a template switch to the 3' end of the minus strand template. This would require the dissociation of the 9 base \([r]\) sequence and base-pairing of the plus strand \([r']\) sequence to the \([r]\) sequence present on the 3' end of the minus strand, thus allowing the continued synthesis of the plus strand, resulting in the formation of the open circular DNA structure present in the Dane particle. Features that may facilitate the dissociation of the \([r/r']\) duplex, could be the presence of the 5' bound protein on the minus strand destabilising the DNA duplex and the high A+T content (hence low \(T_m\)) of the \([r]\) sequence, a feature present in all the Hepadnaviral \([r]\) sequences (fig 1.9). Conservation of the A+T rich sequence would be consistent with the requirement for the transfer of the second strand by a mechanism dependent upon local melting and duplex formation.

The newly synthesised virion DNA could either be transported to the nucleus to complement the pool of covalently closed supercoiled molecules used for transcription of the viral mRNA (Tuttleman et al., 1986), or remain within the core particles and form mature virions surrounded by the surface envelope.
After entry of the virus into the cell, the partially single stranded virion DNA is repaired, resulting in a double-stranded covalently closed DNA molecule present in the nucleus. The covalently closed DNA is transcribed to give the viral transcripts, which are then translated in the cytoplasm to generate the viral polypeptides. The terminally redundant [R], greater than genomic length pregenomic RNA is packaged into immature core particles containing the reverse transcriptase. The minus strand DNA is primed within the DRI sequence by a protein of unknown origin. The minus strand DNA is elongated by reverse transcription with the RNA template being degraded behind the growing point of the minus strand DNA. Minus strand synthesis continues to the end of the RNA template generating a complete minus strand sequence with a 9 base terminal redundancy [r]. Plus strand synthesis is initiated from a RNA primer, probably derived from the 5' end of the pregenomic RNA, hybridised via the homologous DRI (RNA plus sequence) and DR2 (DNA minus sequence). Initiation of plus strand DNA synthesis begins at the 3' boundary of the DR2 sequence and proceeds to the 5' end of the minus strand. The terminal redundancy [r] probably facilitates an intramolecular template switch which allows the plus strand to be elongated generating the virion open circular DNA. Before completion of the DNA plus strand, the immature core particle associates with the surface antigen envelope and is exported from the cell. Alternatively, the newly synthesised DNA is made fully double stranded and transported to the nucleus to supplement the pool of nuclear supercoiled DNA used for viral transcription.
nucleus
cccDNA
HBV

Transcription
Translation
HBcAg, polymerase, (HBsAg, HBxAg?)

packaging

310±2
cap

priming and reverse transcription of minus strand

3108
5' protein primer

degradation of RNA template

3100±2

priming and synthesis of plus strand

2882±1
strong stop

capped RNA primer

formation of open circular DNA, elongation of plus strand

nucleus
Dane particle
1.7 Viral Infection and Hepatocellular Carcinoma

Many independent lines of evidence have related the occurrence of hepatocellular carcinoma (HCC) to chronic hepatitis B virus infection. These include (i) the close geographical correlation between the prevalence of HBV infection and the incidence of HCC (Szmuness, 1978), (ii) the establishment of HCC derived cell lines containing integrated HBV DNA (Alexander et al., 1976; Aden et al., 1979; Koike et al., 1983), (iii) the epidemiological study of Beasley et al., (1981) which has shown that there is a very strong correlation between chronic HBV infection and the development of primary liver carcinoma and (iv) the occurrence of HCC in association with chronic infection of the GSHV, WHV, and DHBV and integrated DNA in their respective hosts (Summers et al., 1978; Ogston et al., 1982; Marion et al., 1984,1986; Rogler and Summers, 1984; Yokosuka et al., 1985; Popper et al., 1987).

Integration of HBV DNA into the cellular genome, while not essential for replication, may be an inevitable consequence of long-term infections. Integrated HBV DNA has been observed in acute hepatitis, asymptomatic carrier hepatitis, chronic hepatitis, cirrhosis and HCC (Shafritz et al., 1981; Brechot et al., 1981,1982; Chen et al., 1982; Kam et al., 1982). In geographical areas, with a high prevalence of both HBsAg carriers and HCC, at least 80% of tumours from HBsAg-positive patients contain integrated HBV DNA sequences (Hino et al., 1984; Chen et al., 1986). It is clear that in the case of HBV-associated HCC, integrated HBV DNA is present in a large proportion of tumours, however, it remains to be determined whether HBV infection and integration play a direct or causal role in the development of neoplasia.

The most studied hepatoma derived cell line PLC/PRF/5 (Alexander et al., 1976) has been shown to contain integrated HBV DNA (Brechot et al., 1980; Chakraborty et al., 1980; Edman et al., 1980; Marion et al., 1980b; Twist et al., 1981). The integrated HBV sequences have been cloned and analysed (Dejean et al., 1983; Koch et al., 1984b; Shaul et al., 1984; Ziemer et al., 1985). The studies show that at least seven fragments of the HBV genome are integrated in PLC/PRF/5 cells. Analysis of the cloned
viral integrations leads to the conclusion that the viral sequences have been rearranged showing deletion, inversion and duplication of the viral sequences following the integration event(s). There appears to be no pattern of site preference for integration within the viral sequences; with the sites of integration-recombination dispersed over the entire viral genome, with some preference for integration within the double stranded portion of the genome (Ziemer et al., 1985). This is in contrast to an earlier report which suggests that the integration events occurred within the single stranded region of the genome (Koshy et al., 1983). The only intact HBV open reading frame is the surface gene (Shaul et al., 1984; Ziemer et al., 1985), which is expressed, with the cells secreting 22 nm HBsAg particles (MacNab et al., 1976; Merion et al., 1979; Skelly et al., 1979). Similar results were obtained for the integrated HBV sequences of the cell line huSP, in which the integrated HBV sequences were found to rearranged, showing an inverted duplication of both viral and cellular sequences at the viral host junctions (Mizusawa et al., 1985).

Analysis of integrated HBV DNA from solid tumour tissue has also shown that integration of viral DNA tends to be random, however, there is evidence that a semi-specific integration may occur in the cohesive overlap region at the DR1 and DR2 sequences (Dejean et al., 1984; Hino et al., 1986; Fowler et al., 1986; Yaginuma et al., 1987a). Evidence for this comes from the results of Yaginuma et al., (1987a) in which one of their clones, isolated from a liver tumour, contained a terminal redundancy at both viral-host junctions, corresponding to the terminal redundancy [r] found on the virion minus strand. This indicates that HBV DNA integration event probably occurred by recombination via the free ends of the viral minus strand.

Although HBV may integrate via a specific viral site(s), the available evidence suggests that integration into the cellular genome occurs at random, in some cases associated with deletion, inversion and translocation of the cellular sequences (Mizusawa et al., 1985; Rogler et al., 1985; Yaginuma et al., 1985,1987a; Hino et al., 1986). Recent reports suggest that integration into the cellular sequences may not be totally
random; highly repetitive sequences have been noted in the cellular flanking regions of the integrations. It has been proposed that chromosomal satellite and mini-satellite repetitive sequences may be "hot spots" for the integration of HBV DNA (Shaul et al., 1984,1986b; Fowler et al., 1986; Berger and Shaul, 1987).

The mechanism by which HBV may induce the development of malignant transformation is not known. One mechanism by which some retroviruses cause neoplastic changes is the expression of viral oncogenes, originally of cellular origin, which disrupt the normal regulation of cell growth. Could a similar mechanism be applied to HBV? Gough and Murray, (1982) transfected Rat2 cells with cloned HBV DNA; the resultant cell line Rat2/130.4/TK4 contains complete integrated copies of the HBV genome, with all the known viral genes transcribed and expressed (Gough and Murray, 1982; Gough 1983; Pugh et al., 1986). Gough and Murray, (1982) could not detect any phenotypic transformation of the cell line Rat2/130.4/TK4, favouring the notion that HBV does not encode an oncogene. However, the results of the above studies should be viewed with caution since the Rat2 cell line is fibroblastoid; the oncogenic functions of HBV may be tissue specific and only be active in infected hepatocytes.

No function of the viral life cycle has been associated with product of the X gene. Therefore, could the product of the X gene have oncogenic functions? Miller and Robinson, (1986) have evaluated the codon usage preference of the hepadnaviral genes. Like retroviral oncogenes, the X gene codon usage preference is similar to that of the genes of eukaryotic cells, with the eukaryotic cellular codon usage preference extending upstream of the viral enhancer. Conversely, the other HBV genes clearly share the codon preference of viral genes. Therefore, Miller and Robinson, (1986) propose that both the HBV enhancer and X gene have been acquired from a cellular origin. Miller and Robinson, (1986) could not detect any significant amino acid homology between the HBV X gene predicted amino acid sequence, and known retroviral oncogenes or cellular proteins. They propose that the HBV X gene could be analogous to the trans-activating proteins of some retroviruses.
Expression of the trans-activating proteins could result in cellular transformation by activation of cellular genes, with tumour formation after a long latent period. The recent results of Wollersheim and Hofschneider, (1987) suggest that the HBV X protein may indeed have trans-activation properties, supporting the proposal of Miller and Robinson, (1986). However, it has not been shown that the trans-activation properties of the HBV X protein are capable of directing the development of malignant transformation.

A second mechanism, whereby viruses cause neoplastic transformation, is "insertional mutagenesis". Disruption of normal transcriptional control may result from deletion, chromosomal translocation or integration of the viral genome, resulting in the introduction of a promoter or enhancer sequence near cellular proto-oncogenes, as has been show for other viruses (Hayward et al., 1981; Klein, 1983). Analysis of cellular flanking sequences of integrated HBV DNA derived from cell lines and human tumour tissues has in most cases failed to detect the presence of any closely linked cellular oncogenes (Dejean et al., 1983; Koch et al., 1984a; Mizusawa et al., 1985). Only in one report has an oncogene-like sequence been found in the cellular flanking sequences. Dejean et al., (1986) report an HBV integration event that has occurred within a human sequence that shares a degree of amino acid homology to the oncogene v-erb-A. Dejean et al., (1986) propose that expression from the HBV surface promoter could result in the synthesis of a preS-v-erb-A "like" fusion protein possibly causing cell transformation. The general absence of oncogenes in close proximity to the integrated HBV DNA does not exclude the possibility of integrated viral sequences activating or augmenting expression of proto-oncogenes placed more distally due to the effect of the HBV enhancer element.

Translocation and deletion of cellular sequences, due to HBV integration (Rogler et al., 1985; Hino et al., 1986), has not been directly shown to result in malignant transformation. However, using one of the HBV-like animal models Moroy et al., (1986) have shown that integrated woodchuck hepatitis virus DNA was associated with enhanced expression and allelic alterations of the c-myc oncogene in three woodchuck hepatocellular
carcinomas. The cloning and analysis of the break-point region of the altered c-myc gene has shown that the c-myc gene was truncated and joined to cellular sequences of unknown function. WHV DNA was not integrated near the c-myc coding exons, excluding the direct role of virus promoters (or possible enhancer) in c-myc activation suggesting that chromosomal translocation has resulted in activation of c-myc expression.

The aetiology of the development of HCC could be multifactorial and HBV infection and integration may play only a casual role in the development of HCC. If the development of HCC is multifactorial, HBV infected cells, or cells containing integrated HBV DNA may be "partially" transformed and have an increased probability of progressing to malignant transformation in association with other environmental factors. Integrated HBV DNA in HCCs may represent merely a "snap-shot" of a previous integration event, which has become detectable through the clonal expansion of the subsequently transformed cell. Whatever the molecular basis of the involvement of HBV in the development of primary liver cancer, it is clear that a period of persistent viral infection, along with chronic liver disease, predisposes to the development of neoplasia.

1.8 The Aims of this Thesis

When this project was initiated in October 1982 very little was known about the molecular biology of the hepatitis B virus. Evidence that HBV infection predisposes to the development of hepatocellular carcinoma was available (Szmuness, 1978; Beasley et al., 1981). The epidemiological studies of Beasley et al., (1981) indicated a correlation between chronic hepatitis B virus infection and the development of hepatocellular carcinoma. At this time, at least two hepatoma derived cell lines were known to contain integrated HBV DNA and express the HBsAg (Aden et al., 1979; MacNab et al., 1976). In addition, the study of HBV infected liver samples indicated that long-term infection with HBV is associated with
integration of virus specific sequences into cellular DNA (Brechot et al., 1981; Shafritz et al., 1981; Chen et al., 1982; Kam et al., 1982). The above studies indicate that there is a strong correlation between persistent viral infection, HBV associated cirrhosis, viral DNA integration and the development of hepatocellular carcinoma.

From the studies of Gough and Murray, (1982), in which they constructed a series of rodent cell lines containing integrated HBV DNA, we knew that cells, other than infected hepatocytes, could express the known viral antigens. A later study of Gough, (1983) reported that the transfected rodent cells expressed four viral specific transcripts. The availability of the cell lines of Gough and Murray, (1982) offered the opportunity to investigate if the HBV genomic DNA required specific viral or cellular sequences for integration. The aim of this project was to generate genomic libraries, using the chromosomal DNA of the transfected rodent cells of Gough and Murray, (1982). The chromosomal DNA was cloned into bacteriophage lambda vectors, and clones containing integrated HBV DNA sequences were isolated. The structure of the integrated HBV sequences was determined and the integration sites mapped.

Many of the aims of this thesis have been superseded by the publication of detailed analyses of the natural integration patterns of HBV sequences, found in hepatocellular carcinoma derived tissue culture cells and solid tumour tissues (Ziemer et al., 1985; Mizusawa et al., 1985; Dejean et al., 1984, 1986; Rogler et al., 1985; Hino et al., 1986). The above studies indicate that HBV DNA integration into the host chromosomes is essentially random, with respect to both viral and cellular sequences. The results obtained studying the HBV integration pattern of the transfected rodent cells, is also consistent with random integration of HBV sequences with the integration sites being dispersed over the HBV genome.

Over the last five years considerable advances have been made concerning our understanding of the molecular biology of HBV. The nucleotide sequence of all the major subtypes of HBV have been determined, this has
allowed the identification of the arrangement of the gene sequences and the tentative identification of potential control signals for transcription. The initiation and polyadenylation sites of the major viral transcripts have been mapped, however, apart from the putative promoter sequence proposed by Cattaneo et al., (1983b), no detailed examination of the viral promoters has been reported.

As an extension to the study of the HBV integration pattern, the isolated genomic clones were examined for their ability to act as templates for the viral specific transcripts observed by Gough, (1983). One of the genomic clones was examined for its ability to act as a template for transcription of the major surface mRNA. The results obtained for in vivo expression of HBsAg indicate that the promoter sequences postulated by Cattaneo et al., (1983b), are transcriptionally active in the absence of the preS1 "TATA" box promoter (Rall et al., 1983). The nucleotide sequence of the major surface antigen promoter was found to be contained, at or between, the HBV130 nucleotide positions 1098 to 1264. The major surface antigen promoter sequences correspond to a region of highly conserved nucleotide sequences, found in the surface antigen promoter regions of all the mammalian Hepadnaviruses. The results of transient HBsAg expression assays also suggest that the NF-1 binding site observed by Shaul et al., (1986a) is not essential for expression of the native HBV surface antigen mRNA.
Chapter Two

Materials and Methods

2.1 Bacterial Strains

All *Escherichia coli* strains used are listed below:

5K  
*hsdR* derivative of C600. *F-, thi, thr, leuB6, supE44, tonA21, lacY1, hsdR*($r_{66}, m_{46}$)  *(Hubacek and Glover, 1970)*

5K/R245  
5K containing the plasmid R245 encoding *Tc*, *Eco* RII.  
This study

BHB2688  
N205 *recA*, $\Deltaimm^{34}$, cIts, b2, red3, *Eam4, Sam7*/\lambda  
*(Hohn, 1979)*

BHB2690  
N205 *recA*, $\Deltaimm^{34}$, cIts, b2, red3, *Dam15, Sam7*/\lambda  
*(Hohn, 1979)*

ED8654  
*supE44, supF58, hsdR514, trpR55, lacY1*  *(Barck et al., 1976)*

GM48  
*F-, dam3, dcm6, gal, ara, lac, thr, leu, thi, tonA, tsx*  
*(Marinus, 1973)*

HB101  
*F-, hsdS20*(r$_{66}$, m$_{46}$), *recA13, ara14, proA2, lacY1, galK2, rpsL20, xyE, mtl1, supE44, $\lambda$*  
*(Boyer and Roulland-Dussoix, 1969)*

JM83  
*ara, $\Delta(lac$ proA,B), rpsL, $\varphi$80, lacZ $\Delta$M15*  
*(Vieira and Messing, 1982)*

NM145  
*R245,Tc*, *Eco* RII  *(N. E. Murray, personal communication)*

NM514  
*hsdR, lyc7*  *(Arber et al., 1983)*
NM522  
$hsd\Delta 5$ derivative of 71-18. $hsd\Delta 5, \Delta(lac\ pro)$

/F'  $lac Z \, \Delta M15, \, lac I^q$ (Gough and Murray, 1983)

Q359  
P2 lysogen derivative of Q358. $hsd R(r_m^-, m_m^+), sup F, \phi 80^r, (P2)$. (Karn et al., 1980)

2.2 Plasmids and Bacteriophages

$E.\ coli$ plasmids and bacteriophages used are listed below:

pDWM1  
Contains the mouse hypoxanthine phosphoribosyltransferase (HPRT) mini-gene cloned in pUC8
(Melton et al., 1986)

pHBV130  
$Tc^r$ This plasmid contains a greater than genomic length HBV genome, approximately 3-8 kb [ca. nucleotides 850±50 to 1460±50] inserted into the Pst I site of pBR322 by oligo dG-dC tailing. (Gough and Murray, 1982)

pHBV130.4  
$Tc^r$ This plasmid contains four tandem copies of unit length HBV DNA cloned between two copies of the vector pBR322. (Gough and Murray, 1982)

pUC8  
$E.\ coli$ cloning vehicle containing a portion of the lac $Z$ gene (β-galactosidase) and multiple cloning sites. (Vieira and Messing, 1982)

λEMBL4  
$\lambda$ b189, 'sbD \lambda 3(R,B,S-red ^+, gam ^+, P_\lambda, trp E^+ -S,B,R) sbD 4 ^*, K H 5 4, chi, n i n 5$ (Frischauf et al., 1983)

λCh34  
$\lambda'$ sr \lambda 1 \lambda 3(R,Ss,Xb,H,B-E. coli stuffer 34 (16.2 kb)
- B,H,Xb,Ss,R)sr \lambda 3 ^*, gam ^+, AWL 113, KH 5 4, nin 5
(Loenen and Blattner, 1983)

λNM1149  
B538, $imm^+ 34$ (Eco RI, Hind III) (Murray, 1983)
\( \lambda \nu_2, \nu_1, \nu_3 \) (Ordal and Kaiser, 1973)

M13mp18/ E. coli M13 based vectors used for DNA sequencing templates, M13mp19 these vectors contain a portion of the lac Z gene (\( \beta \)-galactosidase) and multiple cloning sites (Norrander et al., 1983; Yanisch-Perron et al., 1985).

2.3 Suppliers of Laboratory Reagents

Restriction endonucleases were purchased from Amersham International Ltd., Boehringer Mannheim GmbH Biochemica, Bethesda Research Laboratories Inc., New England Biolabs Inc. and Pharmacia Fine Chemicals AB. E. coli DNA polymerase I (holoenzyme and Klenow fragment) were purchased from Boehringer Mannheim GmbH Biochemica. All enzymic reactions were performed in accordance with the specifications of the manufacturer, except where otherwise noted.

Deoxynucleotide triphosphates were purchased from Bethesda Research Laboratories Inc. Dideoxynucleotide triphosphates were purchased from PL Biochemicals Inc. All radiochemicals were purchased from Amersham International Ltd.

Standard laboratory reagents were purchased from BDH Chemicals Ltd., Boehringer Mannheim GmbH Biochemica, Fisons Chemicals and Sigma Chemical Co., except where otherwise noted.
2.4 Bacteriological Media

All media used was sterilised by autoclaving. Where appropriate, antibiotics were added to molten agar and liquid media to final concentrations: Ampicillin 100 μg/ml, Tetracycline 25 μg/ml.

L broth:
Difco Bacto Tryptone, 10 g; Difco Bacto yeast extract, 5 g; NaCl, 5 g; per litre adjusted to pH 7.2.

L agar:
Difco Bacto Tryptone, 10 g; Difco Bacto yeast extract, 5 g; NaCl, 10 g; Difco agar, 15 g; per litre adjusted to pH 7.2.

BBL agar:
Baltimore Biological Laboratories trypticase, 10 g; NaCl, 5 g; Difco agar, 10 g; per litre (pH unaltered).

BBL top agar:
As for BBL agar, but only 6.5 g Difco agar per litre.

BBL top agarose:
As for BBL agar, but containing 6.5 g agarose (Miles) per litre replacing the Difco agar.

Spizizen salts (5x):
(NH₄)₂SO₄, 2 g; K₂HPO₄, 14 g; KH₂HPO₄, 6 g; tri-sodium citrate, 1 g; MgSO₄, 0.2 g; per litre.

Minimal medium:
Difco Bacto agar, 15 g; 200 ml Spizizen salts (5x); 10 ml 20% (w/v) glucose; 0.125 ml 5 mg/ml Vitamin B1; per litre.
Phage buffer:
KH$_2$PO$_4$, 3 g; Na$_2$HPO$_4$ (anhydrous), 7 g; NaCl, 5 g; 10 ml 0.1 M MgSO$_4$; 10 ml 0.01 M CaCl$_2$; 1 ml 1% (w/v) gelatin solution; per litre.

2.5 Isolation of Plasmid and Bacteriophage Lambda DNA

2.5.1 Large Scale Isolation of Plasmid DNA

The method used to isolate plasmid DNA was a modification of the method described by Clewell and Helinski, (1972). An overnight culture of a plasmid carrying E. coli strain was diluted 100-fold into 500 ml of L broth (containing 1% glucose and antibiotics as required) and shaken at 37°C to an optical density at 650 nm of 0.8-1.0, at which time solid chloramphenicol was added to a final concentration of 150 μg/ml and the culture shaken for a further 12 hours at 37°C. The cells were harvested by centrifugation and resuspended in 5 ml sucrose mix (25% (w/v) sucrose, 50 mM Tris-HCl pH 8.0, 10 mM EDTA). After incubation for 5 minutes on ice with 5 mg lysozyme, 1.0 ml of 0.5 M EDTA was added and the cells incubated for a further 10 minutes on ice. After this incubation, 13 ml of lytic mix (1% (v/v) Triton X-100, 50 mM Tris-HCl pH 8.0, 62.5 mM EDTA) was added slowly and the suspension mixed by gentle swirling. The lysed cell suspension was left on ice for 10 minutes and then centrifuged at 15,000 rpm for 10 minutes at 4°C in a Sorvall ss34 fixed angle rotor. The cleared supernatant was recovered and the volume measured. The volume of the supernatant was adjusted to 25 ml; 23.75 g CsCl and 2.5 ml of 5 mg/ml ethidium bromide were added to the solution. The plasmid DNA was banded by equilibrium centrifugation using a Sorvall Ti50 rotor at 38,000 rpm for 48-72 hours at 20°C. The CsCl gradient thus formed was viewed under long wavelength UV light and the plasmid DNA removed through the side of the tube using a 21 gauge syringe. The ethidium bromide was removed by extraction with ice-cold butan-1-ol and the aqueous phase dialysed overnight against 10 mM Tris-HCl pH 8.0, 1 mM EDTA to remove the CsCl salt.
2.5.2 Alkaline-SDS Preparation of Plasmid DNA

The method used was essentially that of Birnboim and Doly, (1979). An overnight culture of *E. coli* was used to fill a 1.5 ml microcentrifuge tube. The cells were recovered by centrifugation for 5 minutes and the supernatant discarded. The cell pellet was resuspended in 100 µl of lysis solution [2 mg/ml lysozyme, 25 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% (w/v) glucose] and left on ice for 5 minutes. To the suspension was added 200 µl of alkaline-SDS solution [0.2 N NaOH, 1% (w/v) sodium dodecyl sulphate] and the suspension left on ice for 5 minutes. To the lysed cell suspension was added 200 µl of 3 M sodium acetate pH 5.2, the suspension was mixed and left on ice for 30 minutes. The lysed cells were centrifuged for 10 minutes at 4°C and the supernatant transferred to a fresh microcentrifuge tube. The tube was then filled with room temperature ethanol, mixed and centrifuged for 5 minutes at room temperature. The supernatant was discarded and the pellet resuspended in 100 µl of 0.1 M sodium acetate pH 5.2, followed by the addition of 250 µl cold ethanol; the solution was mixed and left at -70°C for 15 minutes. The solution was then centrifuged for 10 minutes at 4°C and the supernatant discarded, the pellet dried under vacuum and redissolved in 50 µl 10 mM Tris-HCl pH 8.0, 1 mM EDTA. The plasmid DNA recovered can be digested with most restriction endonucleases, half of this material (ca. 0.5 µg DNA) can be visualised on an agarose gel by ethidium bromide staining.

2.5.3 Preparation of Bacteriophage Lambda DNA

The method used to isolate bacteriophage lambda DNA was a modification of the methods described by Thomas and Abelson (1966) and Yamamoto et al., (1970). A fresh overnight culture of *E. coli* was diluted ten-fold into 200 ml of L broth, supplemented with 10 mM MgSO₄, in a 2 litre flask and shaken at 37°C. When the cells had grown to OD₆₅₀ of 0.5, bacteriophage λ was added to the culture at 4-6×10⁹ phage plaque forming units and the cells again shaken at 37°C. The optical density of the culture increased to ca. OD₆₅₀ 2.0 then decreased to ca. 0.5. 

71
When the optical density reached a minimum value, 0.5 ml of chloroform was added and the cells shaken for a further 10 minutes. At this time 8 g NaCl, DNAase and RNAase (to a final concentration of 1 μg/ml) were added. The culture was then left for 1 hour at room temperature. The cell debris was removed by centrifugation for 10 minutes at 10,000 rpm in a Sorvall GSA rotor, and the supernatant recovered. Next 20 g of polyethylene glycol 6000 (PEG 6000) was dissolved in the phage suspension, and the suspension left at 4°C overnight. The phage-PEG pellet was collected by centrifugation in a Sorvall GSA rotor for 10 minutes at 10,000 rpm. The supernatant was discarded and the phage pellet resuspended in 5 ml of phage buffer. An equal volume of CHCl₃ was added to the phage suspension in a centrifuge tube and mixed thoroughly to form a white emulsion. The PEG-CHCl₃ was separated from the phage suspension by centrifugation for 15 minutes at 5,000 rpm in a Sorvall ss34 rotor and the cleared aqueous phase recovered.

The bacteriophage suspension was carefully layered onto a preformed cesium chloride step gradient, which consisted of 1.5 ml of 1.7 g/ml, 2 ml of 1.5 g/ml and 2 ml of 1.3 g/ml cesium chloride made up in phage buffer, in a 14 ml MSE ultracentrifuge tube. The phage particles were banded in the gradient by centrifugation at 38,000 rpm for 2 hours at 20°C in a MSE 6×14 ml swing out rotor. The bluish band of phage particles present between the 1.3 and 1.5 g/ml cesium chloride steps was removed by withdrawal through the side of the tube using a 21 gauge needle. The recovered phage suspension was then mixed with enough 1.5 g/ml cesium chloride to fill a 5 ml MSE ultracentrifuge tube. The page suspension was re-centrifuged at 33,000 rpm for 24 hours at 4°C in a MSE 6×5 ml swing out rotor. The phage band was removed as described above. The cesium chloride was removed from the phage particles by dialysis at 4°C for 1 hour against a 1000-fold volume of 10 mM Tris-HCl pH 8.0, 1 mM EDTA (TE). The dialysis buffer was replaced and dialysis continued for an additional hour. The phage suspension was extracted twice with an equal volume of phenol, equilibrated in TE. A third extraction was performed using an equal volume of equilibrated phenol:chloroform (1:1 ratio). The phage DNA solution was further dialysed against TE to remove the organic solvents, using several
changes of TE over a 24 hour period. The concentration of the phage DNA was measured at OD_{260}, with OD_{260} of 1.0 being equivalent to 50 μg/ml.

2.6 General Nucleic Acid Manipulation Methods

2.6.1 Phenol Extraction of Nucleic Acid Solutions

Proteins present in DNA and RNA solutions were removed by extraction with phenol (re-distilled and equilibrated with 1 M Tris-HCl pH 8.0). An equal volume of phenol was mixed with the nucleic acid solution, and the two phases mixed by gentle inversion. The phases were separated by centrifugation in either a microcentrifuge or in an MSE bench centrifuge at 3,000 rpm for 5 minutes, depending upon the volume. The upper aqueous phase was recovered and re-extracted with an equal volume of phenol:chloroform:isoamyl alcohol (50:50:1). The phases were separated by centrifugation and the aqueous phase recovered. Trace amounts of the organic solvents present in the aqueous phase were removed by either i) extraction with an equal volume of ice-cold ether, followed by precipitation of nucleic acids with ethanol (see below), or ii) dialysis of the aqueous phase against several changes of a 1000-fold volume of 10 mM Tris-HCl pH 7.5, 1 mM EDTA.

2.6.2 Precipitation of Nucleic Acids

a) Ethanol precipitation

Nucleic acids were precipitated from aqueous solutions by the addition of one tenth volume of 3 M sodium acetate pH 5.2, and 2.5 volumes of absolute ethanol. After incubation at -70°C for 15 minutes or -20°C overnight, nucleic acids were sedimented by centrifugation in a microcentrifuge for 10 minutes or a Sorvall ss34 or HB4 rotor at 4°C for
10 minutes at 10,000 rpm. The supernatant was discarded, the pellet dried under vacuum and redissolved in 10 mM Tris-HCl pH 7.5, 1 mM EDTA.

b) Isopropanol precipitation

Nucleic acids were precipitated from aqueous solutions by the addition of one tenth volume of 3 M sodium acetate pH 6.0, and 0.6 volumes of isopropanol. After incubation at 0°C for 15 minutes the nucleic acids were sedimented by centrifugation in a microcentrifuge for 10 minutes. The supernatant was discarded, the pellet dried under vacuum and redissolved in 10 mM Tris-HCl pH 7.5, 1 mM EDTA.

2.6.3 In Vitro Enzymic Reactions

a) Digestion of DNA with restriction endonucleases

DNA was digested with three to ten fold excess of restriction endonuclease in 1-5 ml microcentrifuge tubes for 1 hour using conditions recommended by the manufacturer. Digestions were terminated by heating at 65°C for 10 minutes and/or extraction with an equal volume of equilibrated phenol.

b) Modification of recessed 3' termini of DNA

The 3' recessed termini created by digestion of DNA with some restriction enzymes, were rendered blunt ended by the action of E. coli DNA polymerase I (Klenow fragment). The reactions were carried out using 1-10 μg of DNA, with 1 unit of Klenow enzyme per 1 μg DNA, in 10 mM Tris-HCl pH 8.0, 10 mM MgCl₂, and 0.25 mM each of dATP, dCTP, dGTP and dTTP at room temperature for 1 hour.
c) *In vitro* recombination of DNA fragments

i) DNA fragments with 5' or 3' cohesive ends

The DNA fragments were ligated using T4 DNA ligase under the following conditions; linearised vector and insert DNA were mixed in a molar ratio of 1:3 to a final concentration of 5-30 µg/ml. The DNA fragments were covalently joined by the action of T4 DNA ligase at 1 unit ligase per 1 µg DNA, in a buffer consisting of 0.66 M Tris-Cl pH 7.2, 10 mM EDTA, 0.1 M MgCl₂, 0.1 M dithiothreitol, 1 mM ATP; at 10-15°C for 12 hours.

ii) DNA fragments with blunt ends

Reactions were carried out in a similar fashion to those with DNA fragments bearing cohesive ends, however, higher concentrations of DNA (200-500 µg/ml), T4 DNA ligase (ca. 10 units/µg) and ATP (10 mM) were used.

2.7 Sucrose Gradient Fractionation of DNA Fragments

Restriction endonuclease generated DNA fragments were fractionated on pre-formed continuous 12 ml sucrose gradients (10% (w/v) to 40% (w/v) sucrose, dissolved in 20 mM Tris-Cl pH 8.0, 5 mM EDTA) prepared in 14 ml MSE ultracentrifuge tubes. The DNA samples (dissolved in 10 mM Tris-Cl pH 8.0, 1 mM EDTA) were carefully layered onto the surface of the sucrose gradients and centrifuged at 30,000 rpm for 16 hours in a 6x14 ml MSE swing-out rotor. Five drop fractions were collected in microcentrifuge tubes by puncturing the bottom of the ultracentrifuge tubes. Fractions were analysed by agarose gel electrophoresis (see below). Fractions containing the desired DNA fragments were pooled and dialysed against several changes of 10 mM Tris-Cl pH 8.0, 1 mM EDTA. The DNA was recovered by ethanol precipitation and resuspended in 10 mM Tris-Cl pH 8.0, 1 mM EDTA.
2.8 Gel Electrophoresis of Nucleic Acids

2.8.1 Agarose Gel Electrophoresis of DNA

DNA was size fractionated on submerged horizontal agarose slab gels. The agarose gels consisted of 0.7% to 1.0% (w/v) agarose (Miles), containing 0.5 µg/ml ethidium bromide in Tris-borate buffer (TBE) [89 mM Tris borate pH 8.2, 89 mM H₂BO₃, 2 mM EDTA]. Prior to loading the DNA samples, 0.25 volumes of DNA sample buffer (5x) [12.5% (w/v) Ficoll, 50 mM EDTA, 0.125% (w/v) bromophenol blue, 0.125% (w/v) xylene cyanol FF; in (5x) TBE] were added and the samples heated to 65°C for 10 minutes. Electrophoresis was carried out in TBE buffer, between 5-20 volts/cm until the required separation was achieved. Migration of DNA fragments was monitored by the use of a hand held UV-illuminator (254 nm) and observing the migration of the DNA bands due to the fluorescence of the ethidium bromide stain. DNA size standards used were restriction digests of λcEB57, 1 kb and 123 base pair ladders (Bethesda Research Laboratories).

2.8.2 Gel Photography

Nucleic acids were visualised in ethidium bromide stained gels using a Cromato-vue UV-transilluminator (302 nm wavelength). Photographs of gels were taken on a Kodak specialist 3 camera containing a red (A1) filter at f/4.5 for 20 seconds using Ilford HP5 film (5x4 inch sheet film). The photographic films were developed at room temperature in Ilford microphen developer (5 minutes), transferred to 3% acetic acid (30 seconds), fixed in Hypam fixer (5 minutes), washed in cold water for 15 to 20 minutes and air dried. Relative mobility of DNA fragments were measured directly from the negatives and the sizes of the DNA fragments determined relative to size standards, using the relationship of the log₁₀ molecular size verses the relative mobility.
2.8.3 Electrophoresis of RNA on Agarose Gels

RNA was size fractionated by electrophoresis through gels containing formaldehyde (Lehrach et al., 1977). Agarose gels were prepared as follows: 1-3% (w/v) agarose gels were prepared by melting the required amount of agarose (Miles) in distilled H₂O and cooling the solution to 60°C. One tenth the volume of 10×MOPS buffer [0.2 M 3-(N-Morpholino)propane-sulphonic acid (MOPS), 50 mM sodium acetate pH 7.0, 10 mM EDTA] was added followed by formaldehyde stock solution (12.3 M) to a final concentration of 2.2 M. The molten solution was then poured into the gel formers. The RNA samples were prepared in dH₂O; an equal volume of sample buffer (2×MOPS buffer, 50% (v/v) formamide, 3.5 M formaldehyde) was added and the samples denatured by heating at 65°C for 5 minutes, then chilled on ice. Next 0.25 volumes of loading buffer [30% Ficoll type 400, 0.1 M EDTA, 0.1 mg/ml bromophenol blue] were added and the samples immediately loaded onto the prepared gel. The gels were run submerged in 1×MOPS buffer at 5-20 volts/cm until the desired migration was achieved. Under these conditions the bromophenol blue dye migrated slightly slower than rRNA. Molecular weight standards used were as for DNA agarose gels, except the standard samples were treated in the same fashion as the RNA samples. Mouse 28S rRNA and 18S rRNA were also used as molecular weight standards.

2.8.4 Electrophoresis of RNA on Polyacrylamide Gels

Small RNAs were separated using agarose/acrylamide composite gels containing 2.2 M formaldehyde. These gels consisted of 3.5% (w/v) to 5% (w/v) polyacrylamide and 1.0% (w/v) agarose for textile strength, using the buffer conditions described above. Acrylamide stock solution was 30% (w/v) acrylamide [28.5% (w/v) acrylamide, 1.5% (w/v) N,N'-methylene bisacrylamide]. Gels were prepared in vertical gel plates using, for 100 ml gel volume: 1.0 g agarose (Miles) in H₂O (to a final volume of 100 ml) boiled and then cooled to 60°C, acrylamide stock solution to desired final concentration, 17.9 ml of 12.3 M formaldehyde, 10 ml 10×MOPS buffer, 2.1 ml of 3.0% (w/v) ammonium persulphate and
30 μl TEMED (N,N,N',N'-tetramethylethylenediamine). Electrophoresis was carried out in MOPS buffer as described above.

2.8.5 Staining of Nucleic Acids Separated in Gels Containing Formaldehyde

After electrophoresis tracks containing standard marker nucleic acids were removed and stained in either of the following ways:

a) Ethidium bromide staining

Gels were stained with ethidium bromide (5 μg/ml) in 0.1 M ammonium acetate at room temperature for 1 hour and destained for 2 hours in 0.1 M ammonium acetate. Stained nucleic acids were visualised and photographed as described above for DNA agarose gels.

b) Silver staining of nucleic acids

The method used to silver stain nucleic acids was a modification of the method of Herring et al., (1982). Gels were equilibrated in 0.025 M sodium phosphate pH 6.5, for 1 hour. The nucleic acids were fixed in the gel by immersion in 10% (v/v) ethanol, 0.5% (v/v) acetic acid for 15 minutes. The gels were impregnated with silver nitrate by immersion in 11.2 mM AgNO₃ for 20 minutes with gentle agitation and then rinsed in dH₂O. The gels were immersed in 0.75 M NaOH, 90 mM formaldehyde and the stained nucleic acids allowed to develop for 10 minutes. The reaction was stopped by immersing the gel in 70 mM Na₂CO₃ for 30 minutes. The silver stained gels were stored in sealed plastic bags in the dark.

2.8.6 Recovery of DNA from Low Melting-Temperature Agarose Gels

DNA fragments were recovered from low melting-temperature agarose gels essentially as described by Wieslander, (1979). DNA fragments were separated by electrophoresis in agarose gels as described above.
However, low melting-temperature agarose (Sea Plaque agarose, FMC BioProducts) was used, replacing the normal agarose and the electrophoresis carried out at 4°C. The ethidium bromide stained DNA was visualised using a hand held UV-lamp (254 nm). The desired DNA band(s) was excised from the gel using a sterile scalpel and placed in a 1.5 ml microcentrifuge tube or a 50 ml screw cap Falcon tube, depending upon the volume of gel recovered. An equal volume of 10 mM Tris-HCl pH 7.5, 1 mM EDTA was added to the gel and the tube placed at 65°C until the agarose melted. An equal volume of room temperature equilibrated phenol was added to the DNA solution and the phases mixed by gentle inversion to form a white emulsion. The phases were separated by centrifugation for 10 minutes, either in a microcentrifuge or in a MSE bench centrifuge, at 3000 rpm; a dense white layer of agarose formed at the interface. The upper aqueous phase was recovered and re-extracted with phenol and then extracted with phenol:chloroform:isoamyl alcohol (50:50:1). The aqueous phase was recovered and extracted with butan-1-ol to reduce the volume to ca. 100 to 200 μl (and to remove trace amounts of phenol). The DNA was recovered by ethanol precipitation, the pellet was dried under vacuum and resuspended in 10 mM Tris-HCl pH 7.5, 1 mM EDTA to the desired concentration. DNA prepared in this manner can be used for digestion with restriction endonucleases, ligation and radio-active labelling.

2.9 Radio-active Labelling of DNA

2.9.1 Nick Translation

The method of nick translation (Maniatis et al., 1975; Rigby et al., 1977) was used to label DNA with [α-32P]dNTP. A typical reaction contained 0.5 to 1.0 μg DNA in a volume of 20 μl. The DNA was labelled using 2x10^-5 μg DNAase I and 0.5 units of E. coli DNA polymerase I in a solution of 50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 80 μM dATP, 80 μM dGTP, 80 μM dTTP, 35 mM β-mercaptoethanol, 5 μg/ml bovine serum albumin and 10 to 50 μCi of [α-32P]dCTP (10 mCi/ml; 3,000 Ci/mmol; Amersham). The
labelling reaction was carried out at 15°C for 90 minutes. The reaction was terminated by increasing the volume to 200 µl with 20 mM NaCl, 20 mM Tris-HCl pH 7-5, 2 mM EDTA, 0.25% (w/v) sodium dodecyl sulphate, containing 50 µg/ml tRNA and then extracted with equilibrated phenol. The products of the reaction were separated from unincorporated \([α-^{32}P]dCTP\) by Sephadex G-50 column chromatography.

2.9.2 Random Priming Using Hexadeoxyribonucleotide Primers

Random hexadeoxyribonucleotide primers (PL Biochemicals) were used to radio-label DNA according to the method of Feinberg and Vogelstein (1983, 1984). Plasmid or bacteriophage DNA was cleaved with appropriate restriction endonucleases and the fragments electrophoretically separated on a 1-0% low gelling temperature agarose gel (Sea Plaque agarose, FMC BioProducts) containing 5 µg/ml ethidium bromide, in TBE buffer. After electrophoresis the desired DNA bands were cleanly excised using a scalpel and placed into a pre-weighed 1.5 ml microcentrifuge tube. Sterile distilled water was added to a ratio of 3 ml H₂O/g of gel. The tube was then placed at 95-100°C for 10 minutes to melt the gel and denature the DNA. The DNA/gel was stored at -20°C in small aliquots. Prior to labelling, an aliquot was re-boiled for 5 minutes and stored at 37°C for 15 minutes.

The labelling reaction was carried out at room temperature by addition of the following reagents in the stated order: H₂O (to a final volume of 50 µl), 10 µl of oligo-labelling buffer (see below), 2 µl of 10 mg/ml bovine serum albumin, DNA in agarose (10-50 ng of DNA in up to 32.5 µl of molten gel), 5 µl of \([α-^{32}P]dCTP\) (3000-4000 Ci/mmol, 10 µCi/µl) and 2 to 5 units of *E. coli* DNA polymerase I (Klenow fragment). The reaction was allowed to proceed for 4 to 6 hours. The reaction was stopped by addition of 200 µl of a solution containing 20 mM NaCl, 20 mM Tris-HCl pH 7-5, 2 mM EDTA, 0.25% sodium dodecyl sulphate, 1 mM dCTP. Purification of labelled DNA from unincorporated \([α-^{32}P]dCTP\) was not usually necessary prior to hybridisation since 80-100% of the label was routinely incorporated into the DNA.
Oligo-labelling buffer was prepared as follows:

Solution 0: 1.25 M Tris-HCl pH 8.0, 0.125 M MgCl₂.

Solution A: 1.0 ml of solution 0, 18 μl of β-mercaptoethanol (14 M), 5 μl of 0.1 M dATP, 5 μl of 0.1 M dGTP, 5 μl of 0.1 M dTTP.

Solution B: 2.0 M HEPES (N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid), adjusted to pH 6.6 with 4 M NaOH.

Solution C: Hexadeoxyribonucleotides (PL Biochemicals) suspended at 90 OD units/ml.

Oligo-labelling buffer was prepared by mixing the solution A:B:C at a ratio of 2:5:3. The buffer was stored in 10 μl aliquots at -20°C.

2.9.3 5' DNA Terminus Labelling

T4 polynucleotide kinase was used to label the free 5' ends of DNA by a phosphate exchange reaction similar to the method of Berkner and Folk, (1977). Excess ADP was used to drive the exchange reaction, causing polynucleotide kinase to transfer the terminal 5' phosphate from DNA to ADP. The DNA was re-phosphorylated by transfer of the labelled γ-phosphate of [γ-32P]ATP. The reactions were carried out in a total volume of 25 μl containing ca. 10 pmoles 5' ends of DNA in a solution of 50 mM Imidazole-HCl pH 6.3, 12 mM MgCl₂, 15 mM β-mercaptoethanol, 0.3 mM ADP, 0.5 μM ATP, 0.1 mg/ml bovine serum albumin, plus 50 μCi of [γ-32P]ATP (10 μCi/μl, ca. 6000 Ci/mmole) and 5 to 10 units of T4 polynucleotide kinase. The reaction was incubated at 37°C for 30 minutes. The unincorporated [γ-32P]ATP was separated from the products of the reaction by Sephadex G-50 column chromatography.
2.9.4 Measurement of Radioactivity in Nucleic Acids

a) Precipitation with Trichloroacetic Acid

Five microlitres of the products of a labelling reaction to be assayed were spotted onto two Whatman GF/C glass fibre filter disks. The filters were air-dried. One of the filters was placed in ice-cold 10% (w/v) trichloroacetic acid (TCA) for 15 minutes then washed three times with 20 ml of ice-cold 10% (w/v) TCA followed by 10 ml of ice-cold absolute ethanol. The filters were dried under a heat lamp, and the relative incorporation determined by counting in a toluene-based scintillation fluid (6g/litre butyl-PBD in toluene).

b) Absorption to DE81 paper

Five microlitres of the products of the labelling reaction to be assayed were spotted onto two pieces of Whatman DE81 paper. One piece of paper was air-dried and then washed six times for 5 minutes per wash in 0.5 M Na$_2$HPO$_4$. The piece of paper was then washed twice in water for one minute per wash and once in absolute ethanol for 2 minutes. Both pieces of paper were dried under a heat lamp and the relative incorporation determined by counting in a toluene-based scintillation fluid (6g/litre butyl-PBD in toluene).

2.9.5 Sephadex Gel Filtration

Sephadex G-50 (Pharmacia) was prepared as follows: dry Sephadex G-50 (medium) was added to 10 mM Tris-HCl pH 7.5, 1 mM EDTA, and the powder allowed to absorb the aqueous solution overnight at 4°C. Sephadex G-50 columns were prepared in 10 ml plastic disposable pipettes, plugged with glass wool, and washed with several changes of 10 mM Tris-HCl pH 7.5, 1 mM EDTA. The labelled DNA samples were applied to the top of the column in ca. 200 µl volume, with a reservoir of 10 mM Tris-HCl pH 7.5, 1 mM EDTA connected to the top of the column. The passage of radio-

82
active DNA and free nucleotides was monitored using a mini-monitor Geiger counter. The first radio-active peak to pass through the column was collected. The total incorporated label was determined by either TCA precipitation or absorption to DE81 paper as described above.

2.10 Transformation of Escherichia coli

The methods used were essentially those of Mandel and Higa, (1970) and Lederberg and Cohen, (1974). An overnight culture of Escherichia coli was diluted 50-fold into 50 ml of L broth; the cells were grown at 37°C to an OD₆₆₀ of 0.2-0.4. The culture was then placed on ice for 20 minutes; the cells were then recovered by centrifugation and resuspended in 25 ml of ice-cold 0.1 M MgCl₂. The cells were immediately repelleted and resuspended in 2.5 ml of ice-cold 10 mM CaCl₂. The cells were then stored on ice for at least 30 minutes. The transforming DNA (up to 50 ng) was mixed with 200 μl of the competent cell preparation and left on ice for 30 minutes. The cell-DNA mixture was placed in a water bath at 42°C for 2 minutes and then returned to ice for a further 30 minutes.

For bacteriophage transfections of E. coli, 10-fold dilutions of the DNA/recipient cell mixture were made prior to plating. For bacteriophage λ transfections a 0.1 ml aliquot of transformed-cell dilution was mixed with 0.1 ml of freshly grown E. coli cells in 2.5 ml of molten BBL top agar supplemented with 10 mM MgSO₄, and plated onto BBL agar plates. For M13-λacZ based vectors, the transformed-cell dilutions were mixed with 0.1 ml of plating cells in 2.5 ml molten BBL top agar containing, 30 μl of 20 mg/ml XG (5-bromo-4-chloro-3-indoly-β-D-galactoside dissolved in dimethylformamide) and 20 μl of 20 mg/ml IPTG (isopropyl-thio-β-D-galactoside dissolved in dimethylformamide) and plated onto either BBL or minimal agar plates.

For plasmid transformation of E. coli, the DNA/competent cell mixture was diluted into 1.0 ml of L broth and incubated at 37°C for 1 hour. The
cells were then diluted in L broth and plated onto selective L agar media.

2.11 In Vitro Packaging of Bacteriophage Lambda DNA

The method used for the preparation of in vitro packaging mixes was a modification of those described by Becker and Gold, (1975) and Hohn and Murray, (1977). Using this procedure efficiencies of in vitro packaging of uncut λcl857 DNA were ca. 1×10⁶ plaque forming units/μg (pfu/μg).

The packaging mix consisted of four components:

i) Buffer A

20 mM Tris-HCl pH 8.0, 3 mM MgCl₂, 10 mM β-mercaptoethanol, 1 mM EDTA pH 7.0

ii) Buffer M1

Constituents added in the following order:
110 μl H₂O, 1 μl β-mercaptoethanol, 6 μl 0.5 M Tris-HCl pH 7.4, 300 μl of a mixture of 50 mM spermidine and 100 mM putrescine adjusted to pH 7.0 with Tris base, 9 μl 1 M MgCl₂, 75 μl 100 mM ATP.

iii) Freeze-thaw lysate (FTL)

Four lawns of E. coli strain BHB2688 were grown on L agar plates, three of which were incubated overnight at 30°C while the fourth was incubated overnight at 42°C. The plate grown at 42°C was checked for no growth of bacterial colonies, indicating induction of the lambda lysogen.

The cells from the plates grown at 30°C were resuspended in 10 ml of L broth and used to inoculate three 500 ml volumes of L broth in 2 litre flasks to an OD₆₀₀ of 0.08 to 0.1. The cultures were aerated at 30°C until the optical density rose to OD₆₀₀ of 0.3, at which time the contents of the three flasks were combined in a single 5 litre flask.
The culture was induced by slowly shaking in a 45°C water bath for 15 minutes. The induced culture was then grown at 37°C for three hours with vigorous aeration. The culture was checked for induction by adding 2 drops of CHCl₃ to a 1 ml sample of the culture; if induced, the cell suspension cleared within a few minutes.

After induction the culture was chilled in ice-water and the cells recovered by centrifugation, with 200 ml of culture per centrifuge bottle. The supernatant was discarded and the bottles inverted, on ice, to allow the remaining media to drain from the cells, which was then removed with a pasteur pipette. The cells were kept on ice while resuspending the pellet from each bottle in 0.3 ml of ice-cold 10% (w/v) sucrose, 50 mM Tris-HCl pH 7.4. The semi-liquid cell paste was then dispensed into ice-cold 5 ml centrifuge tubes, with approximately 0.5 ml per tube. To each tube was added 30 µl of fresh lysozyme solution (2 mg/ml in 0.25 M Tris-HCl pH 7.4), followed by gentle mixing and immediate freezing in liquid nitrogen. The cells were allowed to thaw slowly on ice until semi-liquid and viscous; 125 µl of buffer M1 was added to each tube, followed by gentle mixing. The samples were immediately centrifuged at 35,000 rpm for 25 minutes at 2°C using an MSE 6x14 ml swing-out rotor. The supernatant was recovered and dispensed in 100 µl aliquots into pre-cooled microcentrifuge tubes, which were immediately frozen in liquid nitrogen and stored at -70°C.

iv) Sonicated extract (SE)

Two lawns of E. coli strain BHB2690 were prepared, one incubated overnight at 30°C and the other at 42°C. As with the E. coli BHB2688 the plate grown at 42°C should show no growth.

The cells were grown and induced as with preparation of the FTL. However, only one 500 ml culture was inoculated and incubation after induction was for 90 minutes. The cells were pelleted, as for FTL, and resuspended in a total of 4.6 ml buffer A and then lysed by sonication, without foaming, until the suspension was no longer viscous. The cell debris was pelleted by centrifugation for 6 minutes at 6,000 rpm using a
Sorvall ss34 rotor. The supernatant was recovered and dispensed in 50 μl aliquots into pre-cooled microcentrifuge tubes which were then frozen in liquid nitrogen and stored at -70°C.

The packaging reactions were carried out by the addition of the reagents in the following order:

- 7 μl buffer A
- a maximum of 200 ng of λ DNA suspended in 1-2 μl
- 1 μl buffer M1
- 6 μl SE
- 10 μl FTL.

The mixture was incubated at 25°C for 60 minutes and the reaction terminated by the addition of 0.5 ml of phage buffer. Appropriate dilutions of the packaged λ DNA were made in phage buffer and mixed with 0.1 ml of freshly grown plating cells and 2.5 ml of molten BBL top agar (or BBL agarose), supplemented with 10 mM MgSO₄, and plated onto BBL agar plates.

2.12 Nucleic Acid Hybridisation

2.12.1 Plaque Detection by Blotting

The method used to detect recombinant clones using bacteriophage vectors was that of Benton and Davis, (1977). Bacteriophages were plated onto E. coli lawns using BBL top agarose onto dry BBL agar in Petri dishes and grown overnight at 37°C. The Petri dishes were placed at 4°C for 1 hour to promote greater adhesion of the agarose top layer to the agar base. Nitrocellulose (Schleicher and Schuell) filter disks were cut to the same dimensions as the Petri dishes. The nitrocellulose disks were carefully placed over the surface of the plates and allowed to become wet. After 1 minute the disks were removed and placed, plaque side up, on a pad of blotting paper saturated with 0.5 M NaOH, 1.5 M
NaCl, to denature the DNA, and left for 2-5 minutes. The disks were placed in 200 ml of 0.5 M Tris-HCl pH 7.4, 3 M NaCl for 1 minute, then rinsed in 2xSSC (0.3 M NaCl, 0.03 M tri-sodium citrate), dried at 37°C on blotting paper and baked in a vacuum oven at 80°C for 90 minutes. The disks were then used for DNA hybridisation (see below).

2.12.2 Transfer of DNA to Membranes (Southern blotting)

The method used to transfer DNA from agarose gels to nitrocellulose or nylon membranes was essentially that of Southern, (1975). After electrophoresis and photography the agarose gels were trimmed of excess agarose to minimise the amount of membrane required for transfer. The gel was placed in 5-10 volumes of 0.25 M HCl for 15 minutes to partially depurinate the DNA (Wahl et al., 1979). The gel was then rinsed in distilled H₂O and placed in 5-10 volumes denaturing solution (1.5 M NaCl, 0.5 M NaOH) with gentle agitation for 30 minutes. The gel was rinsed in dH₂O and then placed in 5-10 volumes of neutralising solution (1.5 M NaCl, 0.5 M Tris-HCl pH 7.2, 1 mM EDTA) with gentle agitation for 30 minutes.

The gel was placed on a sheet of blotting paper (resting on a solid support) saturated with 20xSSC (3 M NaCl, 0.3 M tri-sodium citrate), with the ends of the blotting paper overhanging the solid support such that they are dipping in a reservoir of 20xSSC, so that they act as wicks. The gel was surrounded with strips of Parafilm, flush with the edges of the gel. A piece of nitrocellulose (0.45 μm pore size) (Schleicher and Schuell) or nylon Hybond-N membrane (Amersham), was cut to the same dimensions as the gel. The membrane was soaked in 2xSSC and carefully placed on the surface of the gel making sure that it did not overhang the edges of the gel and that no air bubbles were trapped between the gel and the membrane. Two sheets of blotting paper (the same dimension as the gel) were soaked in 2xSSC and placed on top of the membrane. A stack of dry blotting paper (5-10 cm high) was placed on top of the wetted blotting paper. A glass plate was placed on top and weighted with a 1-1.5 kg weight to compress the whole sandwich. DNA transfer
was allowed to proceed for 16 hours. The blotting paper was removed and the membrane washed in 2xSSC to remove any adhering agarose; the membrane was then dried at 37°C. Nitrocellulose membranes were then baked in a vacuum oven at 80°C for 90-120 minutes. Hybond-N membranes were wrapped in cling-film and placed, DNA side down, on a UV-transilluminator for 2-5 minutes.

2.12.3 Transfer of RNA to Membranes (Northern blotting)

The method used to transfer RNA from agarose gels containing formaldehyde to solid support membranes was essentially the same as the transfer of DNA (Thomas, 1980). However, the gels were not treated with acid, denaturing solution or neutralising solution prior to transfer. After electrophoresis the gels were placed directly on the transfer apparatus containing 20xSSPE [3-6 M NaCl, 0.2 M sodium phosphate pH 7.7, 20 mM EDTA], replacing the 20xSSC. The transfer allowed to proceed for at least 12 hours and the membranes treated as above. However, the filters were not rinsed in 2xSSC before the RNA was bound to the membranes.

2.12.4 DNA Blot Hybridisation

After DNA has been transferred to nitrocellulose or nylon membranes, using either plaque or Southern blotting, the membranes were pre-hybridised by treating in hybridisation solution, with ca. 5 ml of solution per 100 cm² of membrane. Hybridisation solution consisted of 50% formamide, 6xSSC, 0.5% (w/v) sodium dodecyl sulphate (SDS), 5xDenhardt's solution [1xDenhardt's solution: 0.02% (w/v) bovine serum albumin, 0.02% (w/v) Ficoll (mol. wt. 400,000), 0.02% (w/v) polyvinyl pyroldidone (mol. wt. 360,000)] (Denhardt, 1966). Denatured non-homologous DNA was prepared; 0.5 ml of 1 mg/ml sonicated salmon sperm DNA was heated at 95-100°C for 10 minutes and added to the pre-hybridisation solution. Pre-hybridisation was carried out with shaking, in heat-sealed plastic bags, for at least 1 hour at 37°C.
Hybridisation was carried out by replacing the pre-hybridisation solution with fresh hybridisation solution. Denatured non-homologous DNA was prepared as above, however, also containing a denatured radio-actively labelled DNA probe (1×10⁶–1×10⁸ cpm/µg). Hybridisation was performed with shaking for at least 12 hours at 37°C.

Filters were washed for 15 minutes at room temperature in two changes of 200 ml of 2×SSC, 0.5% (w/v) SDS. The wash solution was replaced with 500 ml of 1×SSC, 0.1% (w/v) SDS and the filters washed at 37°C for 1 hour. A high stringency wash was performed by incubating the filters in 500 ml of 0.1×SSC, 0.1% (w/v) SDS at 65°C for 10–30 minutes. The washed filters were blotted to remove excess fluid, however, the filters were not allowed to dry completely. The filters were then sealed in plastic bags and exposed to X-ray film.

2.12.5 RNA Blot Hybridisation

After RNA had been transferred from agarose gels containing formaldehyde, the nitrocellulose or nylon membranes were treated with RNA hybridisation solution with ca. 5 ml of solution per 100 cm² of membrane. RNA hybridisation solution consisted of 50% (v/v) formamide, 5×SSPE, 5×Denhardt's solution, 10% (w/v) dextran sulphate. Denatured non-homologous DNA was added as described above. Pre-hybridisation was carried out with shaking in heat sealed plastic bags for 4 hours at 42°C.

The pre-hybridisation solution was replaced with fresh hybridisation solution containing denatured heterologous DNA plus denatured radio-actively labelled complementary DNA probe. Hybridisation was performed with shaking for at least 12 hours at 42°C. The membranes were washed with two changes of 5×SSPE for 15 minutes at 42°C. The washing buffer was replaced with 1×SSPE, 0.1% (w/v) SDS and the membranes incubated for 30 minutes at 42°C. A high stringency wash was performed by incubating the membranes in 0.1×SSPE, 0.1% (w/v) SDS for 15 minutes at room
temperature. The membranes were blotted to remove excess fluid, sealed in plastic bags and exposed to X-ray film.

2.12.6 Autoradiography

Autoradiography was performed using Cronex 4 or Kodak XAR-5 X-ray films in cassettes. For $^{32}$P or $^{125}$I labels, the Cronex 4 X-ray films were pre-flashed and exposed at -70°C using intensifying screens (Dupont Cronex Lightning Plus). Kodak XAR-5 X-ray films were exposed directly at -70°C using intensifying screens without pre-flashing. For $^{35}$S labelled DNA sequencing gels, the gels were dried and the X-ray film exposed at room temperature. For $^{35}$S labelled proteins, the gels were impregnated with a scintillator (sodium salicylate) and dried, the X-ray films were pre-flashed and exposed at -70°C. The X-ray films were developed using an automatic film processor.

2.13 Electron Microscopy of DNA Heteroduplexes

The method used for preparation of heteroduplexes was a modification of the method of Davis et al., (1971). The two DNAs, at equal concentration (about 3 μg/ml), were denatured in 50 μl of 0·1 M NaOH, 20 mM EDTA and incubated at 27°C for 10 minutes. The solution was neutralised by the addition of 5·6 μl of 2·0 M Tris base, 1·6 M HCl. Renaturation was achieved by adding 50 μl of deionised formamide (99%) such that final concentration of the solution was 0·1 M Tris pH 8·5, 0·01 M EDTA, 50% (v/v) formamide. The DNA was renatured by incubating the solution at 27°C for 2 hours to allow approximately 50% renaturation. The renatured DNA solution (10 μl) was diluted 20-fold into spreading solution (hyperphase solution) with final concentrations of 0·1 M Tris pH 8·5, 10 mM EDTA, 50% (v/v) formamide, 0·1 mg/ml of cytochrome c and 0·15 μg/ml of DNA. The hyperphase solution was spread onto the surface of the hypophase solution (10 mM Tris-HCl pH 8·5, 1 mM EDTA, 15% (v/v) formamide) by slowly pouring the hyperphase solution down a silica ramp (pre-wetted with hypophase solution). The film of hyperphase solution
was allowed to stand for 1 minute before being picked up. The film was picked up onto a 400 mesh copper grid by touching a parlodion coated grid onto the surface of the film at a distance of about 1 grid width from the ramp-solution boundary. The grid was stained by immersion in uranyl stain [50 \(\mu\text{M}\) uranyl acetate, 50 \(\mu\text{M}\) HCl in 90\% (v/v) ethanol] for 30 seconds with gentle agitation and then allowed to dry on filter paper. The grids were rotary shadowed with platinum and coated with a thin layer of carbon. The parlodion was removed by immersion in absolute ethanol for 60 seconds.

The grids were viewed and electron micrographs prepared using a Siemens 1A electron microscope at 60 to 80 kV accelerating voltage. The contour lengths of the heteroduplexes were measured by projecting an image of the electron micrograph onto a white surface and measured using a Keuffel and Esser No. 62 0300 map measurer connected to a digitiser (Ferranti Cetec Graphics) and programmable calculator (British Olivetti Ltd.). The map lengths of the heteroduplexes were measured by comparison to molecular size standards spread with the heteroduplexes (double stranded standard: pAT153, 3657 base pairs; single stranded standard: wild type M13 bacteriophage DNA, 6407 nucleotides).

2.14 DNA Sequencing

The method used to sequence DNA was based upon the methods of Sanger et al., (1977,1980).

2.14.1 Preparation of M13 Replicative Forms

Double stranded M13 vector DNA (RF DNA) was prepared as follows: Appropriate dilutions of the M13 vector bacteriophages were plated on lawns of *E. coli* NM522, prepared in 2.5 ml of molten BBL top agar, containing 30 \(\mu\text{l}\) of 20 mg/ml XG (5-bromo-4-chloro-3-indoly-\(\beta\)-D-galactoside dissolved in dimethylformamide) and 20 \(\mu\text{l}\) of 20 mg/ml IPTG (isopropyl-thio-\(\beta\)-D-galactoside dissolved in dimethylformamide) on minimal agar plates supplemented with vitamin B1. A single blue plaque
was picked with a sterile toothpick and transferred into 2.5 ml L broth and shaken at 37°C for 6 hours. After the incubation, 1.0 ml of the phage/cell culture was added to 2 x 250 ml cultures of *E. coli* NM522 at $OD_{600}=0.25$ and grown for a further 5 hours at 37°C with constant agitation. The M13 infected cells were recovered by centrifugation at 10,000 rpm for 10 minutes in a Sorvall GSA rotor. The cells were lysed and the double-stranded covalently closed circular replicative forms of the M13 vectors were isolated as described for large scale isolation of *E. coli* plasmids.

### 2.14.2 Cloning into the M13 Vectors

The M13-based vectors used were mpl8 and mp19. M13 vector RF DNA and the DNA to be cloned were digested with appropriate restriction endonucleases and ligated *in vitro*. The ligated DNA mixture was used to transform *E. coli* NM522; 0.1 ml of 10-fold serial dilutions (in L broth) of the transformed cells were mixed with 0.1 ml of freshly grown plating cells and 2.5 ml of molten BBL top agar supplemented with 30 µl of 20 mg/ml XG and 20 µl of 20 mg/ml IPTG and plated onto minimal agar containing vitamin B1 and grown at 37°C overnight. The recombinant clear ("white" non-blue) plaques were picked and transferred in a grid formation onto two duplicate minimal agar plates freshly overlaid with a lawn of *E. coli* NM522, and incubated at 37°C overnight. One of the replica plates was used for plaque blotting and hybridisation as described above. Plaques giving positive signals were picked from the remaining duplicate plate, diluted in phage buffer and titrated on *E. coli* as described above; the plates were incubated at 37°C overnight and then stored at 4°C. Recombinant M13 bacteriophages were also stored in suspension by picking a isolated "white" plaque into 50 µl of 20 mM Tris-HCl pH 7.9, 20 mM NaCl, 1 mM EDTA and stored at 4°C.
2.14.3 Preparation of Single-Strand Template DNA

Single-stranded template DNA was prepared by picking an isolated "white" plaque into a 1-5 ml culture of *E. coli* NM522 (an overnight culture freshly diluted 1/100 in L-broth). The inoculated culture was grown at 37°C with vigorous agitation for 4-5 hours. The bacteriophage particles were isolated as follows: the 1-5 ml culture was transferred to a microcentrifuge tube and the cells removed by centrifugation for 10 minutes. The supernatant was transferred to a fresh microcentrifuge tube and the bacteriophage particles precipitated by the addition of 200 µl of 2.5 M NaCl, 20% (w/v) polyethylene glycol 6000 (PEG 6000). The contents of the tube were mixed by inversion and the tube left at room temperature for 30 minutes. The bacteriophage/PEG precipitate was isolated from the solution by centrifugation for 5 minutes; the supernatant was discarded and the tubes inverted onto absorbent paper to allow the remaining medium to drain away. Any remaining solution was removed using a drawn out Pasteur pipette. The pellet was resuspended in 100 µl of sequencing TE (10 mM Tris-HCl pH 7.4, 0.1 mM EDTA). The bacteriophage suspension was extracted with 50 µl of equilibrated phenol and the phases were mixed vigorously for 10 seconds; the tubes were allowed to stand for 10 minutes at room temperature and then re-mixed. The phases were separated by centrifugation for 5 minutes and the aqueous phase recovered, being careful not to take any phenol. The single-strand DNA was precipitated by the addition of 10 µl of 3 M sodium acetate pH 5.2 and 250 µl of absolute ethanol, mixed and then placed at -70°C for 15 minutes. The DNA was recovered by centrifugation for 10 minutes at 4°C and the supernatant discarded. The pellet was rinsed by adding 1 ml of 95% (v/v) ethanol (-20°C), re-centrifuged for 5 minutes. The pellet was dried under vacuum and resuspended in 40 µl of sequencing TE and stored at -20°C.
2.14.4 DNA Sequencing Reactions

i) Annealing reaction

Single-stranded template was annealed to a synthetic M13 sequencing primer (-20, 17 mer; New England Biolabs, M13 sequencing primer #1211). Annealing reactions were performed in 1-5 ml microcentrifuge tubes; 8 μl of single stranded template preparation (see above) was mixed with 1 μl of TM buffer [0.1 mM Tris-HCl pH 7.5, 0.1 mM MgCl₂] plus 1 μl of primer (0.5 pmol/μl). The mixture was heated to 60°C for 1 hour and then cooled to room temperature.

ii) Synthesis reaction

DNA sequencing reactions were carried out in 1-5 ml un-capped microcentrifuge tubes as follows: 2 μl aliquots of the annealing mixture were dispensed into 4 separate microcentrifuge tubes followed by 2 μl of one of the nucleotide mixes (G* or A* or T* or C*, see below). Finally 2 μl of enzyme/label mix [10 mM Tris-HCl pH 7.5, 0.5 μCi/μl [α-35S]dATP (400 Ci/mmoll; Amersham), 0.25 units/μl Klenow fragment, 10 mM dithiothreitol] was added to each tube; the solutions were mixed by centrifugation for 5 seconds and left at room temperature for 25 minutes. After the initial incubation, 2 μl of chase mix [0.25 mM each of dATP, dCTP, dGTP, dTTP] were added to each tube, mixed and incubated at room temperature for another 20 minutes. Prior to polyacrylamide gel electrophoresis, 2 μl of formamide-dye mix [20 mM EDTA, 0.5% (w/v) bromophenol blue, 0.5% (w/v) xylene cyanol FF dissolved in deionised formamide] were added to each tube and mixed by centrifugation for 5 seconds.
Nucleotide mixes were prepared as follows and stored at -20°C:

G*: 2.5 µl of 50 mM dTTP; 2.5 µl of 50 mM dCTP; 12.5 µl of 0.5 mM dGTP; 15 µl of 1 mM ddGTP; 0.5 ml of sequencing TE; 0.5 ml H₂O.

A*: 2.5 µl of 50 mM dTTP; 2.5 µl of 50 mM dCTP; 2.5 µl of 50 mM dGTP; 7.5 µl of 1 mM ddATP; 0.5 ml of sequencing TE; 0.5 ml of H₂O.

T*: 2.5 µl of 50 mM dCTP; 2.5 µl of 50 mM dGTP; 12.5 µl of 0.5 mM dTTP; 30 µl of 10 mM ddTTP; 0.5 ml sequencing TE; 0.5 ml H₂O.

C*: 2.5 µl of 50 mM dTTP; 2.5 µl of 50 mM dGTP; 12.5 µl of 0.5 mM dCTP; 7.5 µl of 10 mM ddCTP; 0.5 ml sequencing TE; 0.5 ml H₂O.

2.14.5 Polyacrylamide Gel Electrophoresis

Separation of the products of the DNA sequencing reactions were performed on vertical gel apparatus with 6% (w/v) acrylamide, 7 M urea gels (dimensions 40cm x 20cm x 0.3mm), cast between two glass plates, with the notched plate coated in a layer of silicon water repellent (Repelcote; Hopkin and Williams Ltd.). Gels were prepared as follows: H₂O to final volume of 40 ml, 6 ml of 40% acrylamide [38% (w/v) acrylamide, 2% (w/v) N,N'-methylene bisacrylamide], deionised using an ion-exchange resin and stored at 4°C, 17 g urea, 4 ml sequencing TBE (10×) [sequencing TBE (1×): 0.90 M Tris borate pH 8.3, 0.90 M H₃BO₄, 25 mM EDTA], 250 µl of 10% (w/v) ammonium persulphate and 40 µl of TEMED. The solution was poured between the glass plates and allowed to polymerise with the well comb in place.

Prior to loading the DNA sequencing reactions/formamide-dye mixture, the samples were boiled for 5 minutes. Approximately 2 µl of each sample was loaded onto the gel using a drawn-out capillary tube. The DNA
sequencing products were separated by electrophoresis in sequencing TBE (1×), at constant power of 25-30 watts (ca. 20-30 mA, 1.1-1.5 kV).

Electrophoresis was continued until the required fragment separation was achieved (bromophenol blue migrates with DNA fragments of about 25-30 nucleotides, the xylene cyanol FF migrates with DNA fragments 75-80 nucleotides). After electrophoresis the glass plates were prised apart making sure that the gel remained attached to one of the glass plates. The gel was immersed in 10% (v/v) acetic acid, 10% (v/v) methanol fixing solution. After 15 minutes in the fixing solution, the gel (still on the glass plate) was removed and the acetic acid/methanol solution was allowed to drain off. The gel was transferred to a sheet of blotting paper, covered in cling-film and dried on a gel slab drier. Once dry the cling-film was removed and the dried gel exposed to X-ray film.

2.15 In Vitro Transcription of Eukaryotic Genes

The method used for in vitro transcription experiments was that of Manley et al., (1980). The in vitro eukaryotic transcription system was purchased from Bethesda Research Laboratories; 25 μl reactions were carried out according to Manley et al., (1980).

A typical reaction (25 μl) consisted of the following reagents: 15 μl of HeLa cell extract (BRL); 0.2-1.5 μg of linear template DNA; 0.5 mM each of ATP, CTP, TTP; 0.05 mM GTP; 0.14 mM EDTA; 1 mM creatine phosphate; 5 μCi [α-32P]GTP (10 μCi/μl, 410 Ci/mmol; Amersham); sterile H2O to final volume of 25 μl. The incubation was carried out at 30°C for 90 minutes.

The reaction was terminated by the addition of 250 μl of 7.5 M urea, 0.5% (w/v) SDS, 10 mM Tris-HCl pH 8.0. An equal volume (275 μl) of phenol:chloroform:isoamyl alcohol (20:20:1) was added and the phases mixed. The phases were separated in a microcentrifuge for 5 minutes. The aqueous phase (first aqueous phase) was transferred to a fresh microcentrifuge tube, leaving the organic phase plus interphase. To the organic phase was added 150 μl of 7 M urea, 0.35 M NaCl, 1% (w/v) SDS,
10 mM EDTA, 10 mM Tris-HCl pH 8.0 plus 25 μg of yeast tRNA. The phases were mixed by vortexing and then separated in a microcentrifuge for 20 seconds. A very dense protein precipitate formed at the interphase, therefore, the organic phase was removed leaving the interphase. The interphase plus second aqueous phase were re-extracted with 200 μl of chloroform; after vortexing and centrifugation (20 seconds) the second aqueous phase was recovered and combined with the first aqueous phase. The combined aqueous phases were extracted with an equal volume of phenol:chloroform:isoamyl alcohol and the phases separated by centrifugation for 5 minutes and the aqueous phase recovered. The aqueous phase was re-extracted twice with an equal volume of chloroform. The nucleic acids precipitated from the aqueous phase by adding 1 ml of ethanol, mixing and incubating in a dry-ice ethanol bath for 5 minutes. The nucleic acids were recovered by centrifugation for 10 minutes at 4°C. The supernatant was discarded and the pellet resuspended in 200 μl of 0.2% (w/v) SDS and 200 μl of 2 M ammonium acetate and then precipitated with ethanol as described above. The resuspension and precipitation was repeated and the nucleic acid pellet dried under vacuum, and dissolved in sterile distilled H₂O.

The labelled RNA samples were separated by electrophoresis on either an agarose or polyacrylamide/agarose composite gels containing 2-2 M formaldehyde in MOPS buffer as described previously. Upon completion of electrophoresis, the gels were dried using a slab gel drier and autoradiographed at -70°C in contact with Kodak XAR-5 X-ray film and an intensifying screen. Molecular weight standards used were ³²P-labelled DNA fragments visualised by autoradiography. Mouse (28S, 18S, 5.8S and 5S) and E. coli (5S) ribosomal RNAs were also used and visualised either by ethidium bromide staining or silver staining of non-dried portions of the gels.
2.16 Eukaryotic Cell Culture

2.16.1 Cell Lines and Growth Conditions

Tissue culture cells lines used were: Rat2 (tk-, derivative of the rat fibroblastoid cell line Rat1) (Topp, 1981), Mouse LMtk- cells (Kit et al., 1963), the HPRT-deficient Chinese hamster cell line RJK88 (Fuscoe et al., 1983) and derivatives of the above cell lines (Gough and Murray, 1982). Cells were grown in DMEM (Dulbecco's modified Eagles' medium; Gibco) supplemented with 7% (v/v) foetal bovine serum (Flow laboratories) at 37°C in a CO₂ incubator. After transformation (see below) tk+ and HPRT+ cells were maintained in HAT medium (DMEM containing 15 µg/ml hypoxanthine, 0.2 µg/ml aminopterin and 5 µg/ml thymidine).

2.16.2 Transformation of Mammalian Cells with Cloned DNA

a) Calcium phosphate precipitation

The procedure used was a modification of the methods of Graham and van der Eb, (1973) and Wigler et al., (1978). DNA was co-precipitated with calcium phosphate and introduced to tissue culture cells using the following procedure. A known amount of sterile DNA, ethanol precipitated and resuspended in 1 mM Tris-HCl pH 7.8, 0.05 mM EDTA, was added to 1.0 ml HBS (0.14 M NaCl, 5 mM KCl, 0.7 mM Na₂HPO₄, 56 mM D-glucose, 20 mM HEPES pH 7.1). To this mixture was then added 62 µl of 2 M CaCl₂, slowly to the DNA solution with gentle mixing. The solution was then left at room temperature for forty-five minutes to allow a precipitate to form and then added to an aspirated cell culture (at ca. 10-20% confluence in a 100 mm diameter Petri dish). The cell cultures were then left for 20 minutes with frequent gentle mixing. After the incubation, 10 ml of DMEM was added and the cells incubated for four hours at 37°C in a CO₂ incubator. After incubation the medium was removed and the cells exposed to 2 ml of 15% (v/v) glycerol in HBS and incubated for 3-5 minutes. The cells were then washed and incubated
overnight in 15 ml of fresh DMEM. Twenty-four hours after the addition of the DNA the medium was replaced with appropriate selective medium.

b) Transient Assay Procedure

The method used for transient assay transformation was a combination of the methods of Sompayrac and Danna, (1981) and Luthman and Magnusson, (1983). DEAE-dextran and chloroquine were used as facilitators for DNA uptake in a modification of the method of McCutchan and Pagano (1968). Tissue culture cells were plated onto 10 cm diameter Petri dishes (2.5 x 10⁶ cells per dish) and allowed to grow at 37°C in a CO₂ incubator to be subconfluent at the time of transformation. DEAE-dextran (mol. wt. 5 x 10⁵; Sigma) was dissolved in serum-free DMEM, 50 mM Tris-HCl pH 7.3, to final concentration of 200 µg/ml and filter sterilised. The cells were washed twice with serum-free DMEM and incubated with 5 ml DEAE-dextran/DMEM/Tris containing approximately 5 µg of DNA at 37°C in a CO₂ incubator for 6 hours. After the initial incubation the transfection medium was supplemented with chloroquine to a final concentration of 100 µM chloroquine diphosphate (Sigma) and incubated for a further 2 hours at 37°C in a CO₂ incubator. After the incubation the cells were washed once in serum-free DMEM and incubated with DMEM supplemented with 7% (w/v) fetal bovine serum for the required expression time. For HBsAg expression, the cells were allowed to grow to confluence and the medium and cells harvested.

2.16.3 Preparation of Eukaryotic DNA from Tissue Culture Cells

The method used to isolate chromosomal DNA from tissue culture cells was a modification of the method of Gross-Bellard et al., (1973) as described by Gough and Murray (1982). Tissue culture cells were grown to confluent monolayers in 10 cm Petri dishes. The cells were removed from the Petri dishes by treatment with trypsin/EDTA (Gibco), washed with sterile phosphate-buffered saline (PBS) (Dulbecco’s phosphate buffered saline; Flow laboratories) and resuspended in 5.0 ml of 10 mM
Tris-HCl pH 7.5, 10 mM EDTA, 100 mM NaCl per 10 Petri dishes. The cells were lysed by the addition of 250 µl of 10% (w/v) sodium dodecyl sulphate. The lysed cells were incubated with 500 µl of 10 mg/ml proteinase K (Sigma) and 50 µl of 10 mg/ml heat-treated ribonuclease A (Sigma) at room temperature for 16 hours with gentle agitation. The DNA was isolated by extraction, once with phenol:chloroform:isoamyl alcohol (50:50:1) and once with chloroform:isoamyl alcohol (50:1). The aqueous phase was then dialysed for 2 hours against several changes of 10 mM Tris-HCl pH 8.0, 1 mM EDTA. The concentration of DNA was measured by absorbance at 260 nm.

2.16.4 Preparation of Eukaryotic RNA from Tissue Culture Cells

The method used to isolate RNA from tissue culture cells was a modification of the method of Feramisco et al., (1982). Cells, grown to confluent monolayers in 10 cm Petri dishes, were washed with 10 ml of sterile PBS per dish and the cells scraped off each plate using a rubber policeman (in 2 ml of PBS). Using a wide-mouthed pipette the cells were transferred to 50 ml screw cap Falcon tubes. The cells were harvested by centrifugation at 1200 rpm for 5 minutes in a MSE bench centrifuge and resuspended in 2 ml/dish of guanidinium lysis solution [4 M guanidinium thiocyanate (sigma), 50 mM Tris-HCl pH 7.6, 10 mM EDTA, 2% (w/v) Sarkosyl, 0.14 M β-mercaptoethanol]. The cells were lysed and the mixture heated to 60°C. While maintaining temperature, the suspension was drawn through a 18-gauge needle fitted to a 10 ml syringe and the suspension forcefully ejected back into the tube. This step was repeated until the viscosity of the suspension was reduced by shearing of the chromosomal DNA. An equal volume of equilibrated phenol (pre-heated to 60°C) was added to the lysed cells and the emulsion was drawn through the syringe a number of times. A half volume of 0.1 M sodium acetate pH 5.2, 10 mM Tris-HCl pH 7.4, 1 mM EDTA was added to the phenol cell emulsion followed by an equal volume of a 24:1 solution of chloroform:isoamyl alcohol. The suspension was shaken vigorously, while maintaining the temperature at 60°C, for 10-15 minutes. The suspension was cooled on ice and the phases separated by centrifugation.
at 3,000 rpm for 10 minutes in a MSE bench centrifuge. The aqueous phase was recovered and extracted once phenol:chloroform (1:1) and twice with chloroform at room temperature. The aqueous phase was recovered and transferred to sterile corex tubes. The nucleic acids were precipitated by adding two volumes of ethanol and storing at -20°C for at least 2 hours. The nucleic acids were recovered by centrifugation at 10,000 rpm for 20 minutes in a Sorvall ss34 rotor at 4°C. The pellet was dissolved in the original starting volume of 0.1 M Tris-HCl pH 7.4, 50 mM NaCl, 10 mM EDTA, 0.2% (w/v) SDS. Proteinase K was added to a final concentration of 1 mg/ml and the solution incubated for 1-2 hours at 37°C. The nucleic acid solution was extracted once with phenol:chloroform:isoamyl alcohol (25:25:1), once with phenol:chloroform (1:1) and twice with chloroform at room temperature. The nucleic acid was precipitated by adding one tenth of the volume of 3 M sodium acetate and 2.5 volumes of ethanol and placing at -70°C for 15 to 30 minutes. The nucleic acids were recovered by centrifugation at 10,000 rpm for 20 minutes at 4°C in a Sorvall ss34 rotor. The pellet was rinsed with 70% (v/v) ethanol (-20°C), dried under vacuum and resuspended in distilled H₂O.

To select poly(A)^+ RNA (mRNA), the nucleic acids prepared as above were fractionated by chromatography on an oligo-(dT)-cellulose column. Oligo-(dT)-cellulose (Type 7; Pharmacia) was equilibrated in sterile loading buffer [20 mM Tris-HCl pH 7.6, 0.5 M NaCl, 1 mM EDTA, 0.1% (w/v) SDS]. A 1 ml column was poured in a disposable column and washed with three column volumes of i) sterile H₂O; ii) 0.1 M NaOH, 5 mM EDTA; iii) sterile H₂O. The pH of the effluent was checked to be less than 8. The column was then washed with 5 column volumes of loading buffer. The RNA samples (from above) were mixed with an equal volume of (2x)loading buffer and heated to 65°C for 5 minutes, cooled and applied to the column. The flow-through was collected, heated to 65°C, cooled and re-applied to the column. The column was washed with 10 volumes of loading buffer followed by 10 column volumes of 20 mM Tris-HCl pH 7.6, 0.1 M NaCl, 1 mM EDTA, 0.1% (w/v) SDS. The poly(A)^+ RNA was eluted from the column using 3 column volumes of 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.05% (w/v) SDS. The poly(A)^+ RNA was precipitated with ethanol and the
pellet rinsed with 70% (v/v) ethanol (−20°C), dried under vacuum and dissolved in sterile dH₂O.

The column was regenerated by sequential washings with H₂O and NaOH as described above.

2.16.5 Isolation of Total Nucleic Acids from Tissue Culture Cells

The method used to isolate total nucleic acid from tissue culture cells was a modification of the method to isolate cellular RNA. The above procedure was followed, however, the syringe shearing steps were omitted and all extractions were performed at room temperature. The poly(A)+ RNA was not purified from the non-polyadenylated nucleic acids.

2.16.6 In vivo Labelling of Proteins

Tissue culture cells were grown to slightly sub-confluence on 10 cm diameter dishes in DMEM medium. The cell were washed twice in PBS and the medium replaced with 10 ml of MEM (minimum essential medium (Eagle) without L-methionine and L-glutamine; Gibco # 041-1900) supplemented with L-glutamine (584 mg/ml) and 250 μCi of L-[³⁵S]methionine (>800 Ci/mmol; Amersham #SJ.204). The cells were then incubated at 37°C in a CO₂ incubator for 24 hours. The medium was removed and retained at −20°C.
2.17 Protein and Immunological Techniques

2.17.1 Determination of Protein Concentration

Protein concentrations were determined as described by Lowry et al., (1951).

Reagents used were as follows:

Solution A: 2% (w/v) Na$_2$CO$_3$ (anhydrous) in 0.1 M NaOH.

Solution B: 0.5% (w/v) CuSO$_4$.5H$_2$O in 1% (w/v) sodium citrate.

Solution C: 1 ml of solution B plus 50 ml of solution A.

Solution D: Folin and Coicalteau's phenol reagent diluted 1:1 in distilled H$_2$O.

Standards used for protein concentration estimations were bovine serum albumin (BSA) solutions ranging from 0-400 µg/ml BSA in distilled H$_2$O. Protein samples were diluted to 1:50, 1:100, 1:200 and 1:400. Each protein sample (400 µl) was mixed with 2 ml of solution C in a glass 5 ml test tube and incubated for 10 minutes at room temperature; 0.2 ml of solution D was added and the mixture incubated for a further 30 minutes at room temperature. The optical absorption of the standards and samples was measured at a wavelength of 550 nm in a spectrophotometer. The absorption of the samples was related to protein concentration using a standard curve constructed from the absorption values of the known standards.
2.17.2 Radioactive Labelling of Proteins Using Iodogen

i) Preparation of Iodogen coated tubes

Iodogen (Fraker and Speck, 1978) coated 1-5 ml microcentrifuge tubes were prepared as follows: 1 mg of Iodogen (Pierce Chemical Company, # 28600) was dissolved in 25 ml of dichloromethane and distributed into 50 μl aliquots in microcentrifuge tubes. The dichloromethane was evaporated under a stream of nitrogen gas, the tubes were "filled" with N₂ gas, capped and then stored at -70°C.

ii) Iodination reaction

To each tube was added 5 μl of 0.1 M phosphate buffer pH 7.5, 10 μl of protein solution (50-100 μg) and 5 μl of ¹²⁵Iodide (100 mCi/ml; Amersham, # IMS.30). The tubes were incubated at room temperature for 10 minutes with vigorous shaking. The reaction was terminated by transfer of the solution to a fresh 1-5 ml microcentrifuge tube.

The products of the iodination reaction were separated from non-incorporated ¹²⁵I by separation through a Sephadex G-25M column. A PD-10 column (Sephadex G-25M; Pharmacia) was equilibrated with phosphate buffered saline (PBS) containing 1% (w/v) bovine serum albumin. Then 0.5 ml of PBS/0.1% BSA (saturated with potassium iodide) was added to the top of the column. The 20 μl reaction mixture was added to the 0.5 ml of PBS/0.1% BSA/sat. KI and the sample allowed to enter the gel bed. The labelled proteins were then eluted with PBS/1% BSA. Ten to twelve drop fractions (ca. 0.5 ml) were collected and the level of radioactivity present in the fractions determined by counting 5 μl aliquots using a LKB mini-gamma-counter. The fractions corresponding to the first peak eluted from the column were pooled.
2.17.3 Separation of Proteins by Polyacrylamide Gel Electrophoresis

The polyacrylamide gel electrophoresis system used to separate proteins was a sodium dodecyl sulphate-polyacrylamide gel electrophoresis system (SDS-PAGE) using a discontinuous buffer. Polyacrylamide gels were formed between two glass plates with gel dimensions 170mm x 150mm x 1.5mm. The resolving gel (10% polyacrylamide) was prepared by adding the following constituents: 13.24 ml of 30.2% acrylamide solution [30% (w/v) acrylamide, 0.2% (w/v) bis-acrylamide], 5 ml of 3 M Tris-HCl pH 8.8, 0.4 ml of 10% (w/v) SDS, H₂O to a final volume of 40 ml, 0.2 ml of 10% (w/v) ammonium persulphate and 20 μl of TEMED. The solution was poured between the glass plates, leaving a 2-3 cm space for the stacking gel, overlaid with a thin layer of butan-1-ol, to allow a smooth surface to form and left to polymerise for approximately 2 hours. When the resolving gel had polymerised the butan-1-ol was removed and the top edge of the gel washed extensively with distilled water and then allowed to drain. The stacking gel (3.75% polyacrylamide) was prepared by addition of the following constituents: 3.75 ml of 30.2% acrylamide solution, 3.75 ml of 1 M Tris-HCl pH 6.8, 0.3 ml of 10% (w/v) SDS, H₂O to a final volume of 30 ml, 0.2 ml of 10% (w/v) ammonium persulphate and 20 μl of TEMED. The stacking gel solution was poured on top of the resolving gel and the gel comb positioned in place. The stacking gel was allowed to polymerise for approximately 30 minutes. Prior to loading the protein samples, an equal volume of sample buffer [0.12 M Tris-HCl pH 6.8, 20% (v/v) glycerol, 4% (w/v) SDS] was added to each sample; dithiothreitol (DTT) was added to a final concentration of 60 mM DTT. The samples were boiled for 10 minutes before loading onto the gel. Electrophoresis was carried out using protein gel buffer [25 mM Tris, 0.192 M glycine, 0.1% (w/v) SDS, (pH 8.3)] at a constant current of 5-7 mA for 12-16 hours or 25 mA for 5-6 hours until the desired separation was achieved.

Protein molecular weight markers used were low molecular weight markers (Pharmacia) containing: phosphorylase b (94 kd), bovine serum albumin (67 kd), ovalbumin (43 kd), carbonic anhydrase (30 kd), trypsin inhibitor
(20.1 kd) and α-lactalbumin (14.4 kd). A standard curve of log_{10} molecular weight verses distance migrated was constructed using data from the standard polypeptide molecular weight markers.

2.17.4 Staining of Protein Separated on Polyacrylamide Gels

After electrophoresis on polyacrylamide gels, the proteins were stained for 30 minutes in a solution of 0.2% (w/v) Coomassie brilliant blue R (Sigma), 45% (v/v) methanol, 10% (v/v) glacial acetic acid. The gels were destained with gentle agitation in 45% (v/v) methanol, 10% (v/v) glacial acetic acid, until the background staining was reduced to an acceptable level. The gels were stored for short periods in sealed plastic bags or permanently dried onto blotting paper using a slab gel drier.

2.17.5 Western Blotting

Proteins were electrophoretically transferred from non-coomassie blue stained SDS-PAGE gels to nitrocellulose filters (Towbin et al., 1979). The electrotransfer of proteins was accomplished using a Biorad electrotransfer apparatus. A cassette was constructed containing the layered components in the following order: a Scotchbrite pad, 3 sheets of blotting paper cut to the same dimensions as the gel, the gel, a sheet of nitrocellulose (0.45 μm; Schleicher and Schull) the same size as the gel, 3 more sheets of blotting paper and finally another Scotchbrite pad. The cassette was closed and placed in the Biorad electro-transfer tank with the nitrocellulose sheet towards the anode. The electro-transfer was conducted in a buffer of 0.025 M Tris-HCl pH 8.3, 0.15 M glycine, 20% (v/v) methanol, 0.1% (w/v) SDS. Proteins were transferred at a constant current of 0.3-0.5 amperes for 5 hours.

After electro-blotting the nitrocellulose filter was incubated in a solution of 10 mM Tris-HCl pH 8.1, 0.15 M NaCl, 3% (w/v) bovine serum albumin (TS/BSA) for 1 hour, with gentle agitation, to reduce non-
specific binding of protein probes to the membrane. Immunological probing was carried out by gentle agitation at room temperature with the antibody diluted in TS/BSA (5-10 ml solution per 100 cm² of membrane) for 12-16 hours. The filter was washed in 5 changes of 10 mM Tris-HCl pH 8.1, 0.15 M NaCl, 0.1% (v/v) Tween20, over a 60 minute period. After washing the filter was incubated with 5-10 ml, per 100 cm², of 10 mM Tris-HCl pH 8.1, 0.15 M NaCl containing 0.5 μCi/ml of ¹²⁵I-labelled protein A (>30mCi/mg; Amersham, #IM.144) for 90 minutes at room temperature. After incubation the filter was washed as above, wrapped in cling-film and exposed at -70°C using two intensifying screens and pre-flashed X-ray film.

2.17.6 Radioimmunoassays

Radioimmunoassays for the presence of hepatitis B virus surface antigen were performed using an AUSRIA II-125 kit (Abbot laboratories). The method used was procedure B as recommended by the manufacturer. Protein samples (200 μl) were added to wells containing beads coated in guinea pig antibody to the hepatitis B surface antigen. The samples and beads were incubated at room temperature for 16 ± 4 hours. The liquid was removed by aspiration and the beads washed by filling the wells 15 times with distilled H₂O. To each well was added 200 μl of ¹²⁵I-anti-HBS (human) and the beads incubated for 1 hour at 45°C. The liquid was removed by aspiration and the beads washed as described above. The beads were then transferred to counting vials and counted for 1 minute in a LKB mini-gamma-counter.

2.17.7 Double Antibody Immunoprecipitation

Antisera were diluted 1:5 or 1:10 in RIP buffer [phosphate buffered saline (flow laboratories), 0.5% (w/v) bovine serum albumin]. For each assay 50 μl of diluted antiserum was mixed with 100 μl of radio-active protein sample (in PBS). The samples and antiserum were incubated overnight at 4°C with gentle mixing. After the incubation, 100 μl of a 1:2 dilution
(in RIP) of either donkey anti-rabbit serum or goat anti-human serum (Scottish Diagnostics Laboratories) (depending on the antiserum used, either rabbit or human/chimpanzee respectively) was added and the samples incubated overnight at 4°C. After the incubation, 1 ml of RIP buffer, containing 0.1% (w/v) starch, was added and the immuno-precipitate/starch was isolated by centrifugation in a microcentrifuge for 10 minutes. The pellet was washed with 1 ml of RIP buffer and reisolated by centrifugation.

The immuno-precipitate was subjected to electrophoresis using SDS-PAGE. For ¹²⁵I-labelled proteins the gels were dried on a slab gel drier and the gels exposed to X-ray film using indirect autoradiography at -70°C using two intensifying screens and pre-flashed film. For ³⁵S-labelled proteins the gels were impregnated with sodium salicylate and used for fluorography at -70°C with intensifying screens and pre-flashed X-ray film. The method used for impregnating the gel with sodium salicylate was that of Chamberlain, (1979). Coomassie blue stained gels were presoaked for 30 minutes in water to remove acetic acid. The gel was immersed in 1 M (16% w/v) sodium salicylate for 30 minutes. The gel was removed from the sodium salicylate and dried on blotting paper using a slab gel drier.
Chapter Three
Results and Discussion

Cloning and Analysis of Integrated HBV Sequences

3.1 Construction and Screening of Genomic DNA Libraries

3.1.1 Characterisation of Integrated HBV DNA and Cloning Vectors

The cell lines used in this study were those constructed by Gough and Murray, (1982). The cell lines L/130.4/TK154 and Rat2/130.4/TK4 were constructed by co-transfection of mouse L (tk−) and Rat2 (tk−) cells respectively, with the plasmid pHBV130.4 (fig. 3.1) and the thymidine kinase gene from the Herpes simplex virus, contained on the plasmid pFG5 (Colbere-Garapin et al., 1979), as a selectable marker. The resultant cell lines were found to express HBsAg, HBeAg and HBxAg (Gough and Murray, 1982; Pugh et al., 1986).

Examination of the poly(A+) RNA isolated from these cell lines showed that the cells express four viral specific mRNAs of 4.4, 3.9, 2.4 and 1.0 kb in length, which hybridise to the minus strand of HBV, indicating that they were of the plus strand polarity (Gough, 1983). The 2.4 kb mRNA hybridised to the surface and X genes and was designated the surface mRNA in agreement with Chakraborty et al., (1980) and Pourcel et al., (1982). The 1.0 kb mRNA specifically hybridised to the X region of the HBV genome and therefore was designated the X mRNA. The 3.9 and 4.4 kb mRNAs are both greater than genomic length of HBV and were only present in cells that expressed HBCAg/HBeAg. It was proposed that one of these mRNAs codes for the core protein. The 3.9 kb mRNA is of the same order of size as the core mRNA observed in infected liver (Cattaneo et al., 1984) and is probably the same HBV transcript. The 1.0 and 4.4 kb mRNAs have not been observed in infected tissues and may possibly be peculiarities of the expression system. However, transcripts have been observed to be expressed in hepatoma derived cells originating
Figure 3.1. Restriction endonuclease cleavage maps of A) pHBV130 and B) pHBV130·4. The heavy lines represent HBV sequences and the thin lines represent pBR322 sequences. The location and direction of the open translational reading frames are shown on the inner arcs. The polymerase and surface genes of pHBV130 are interrupted by pBR322 sequences. The larger arrow heads on the pHBV130·4 map represent the four head-to-tail genomic length Bst EII fragments of HBV ligated together to form pHBV130·4. Abbreviations: B, Bam HI; Bg, Bgl II; Bs, Bst EII; E, Eco RI; Ps, Pst I; Xh, Xho I. Open reading frames: C, core gene; P, polymerase gene; S, surface gene; preS1 and preS2, preS open reading frames; X, X gene. Reference: Gough and Murray, (1982)
A

pHBV130
8.2 kb

B

pHBV130.4
22.7 kb
immediately upstream of the X gene (Trenin and Laub, 1987; Siddiqui et al., 1987). These transcripts probably correspond to the X mRNA observed by Gough, (1983). It is worth noting that the size difference between the 4.4 and 3.9 kb RNA species corresponds to the difference in the distance of the mapped positions of the observed initiation sites of the core mRNA (Will et al., 1987; Yaginuma et al., 1987b) and the X mRNA (Trenin and Laub, 1987; Siddiqui et al., 1987). Therefore, the 4.4 kb mRNA probably corresponds to transcripts from the X gene promoter which are polyadenylated upon transcription of a complete HBV genome, generating a mRNA with a terminal redundancy of approximately 600 nucleotides.

Examination of the integrated HBV DNA contained within these cell lines (fig. 3.2), indicates that both cell lines contain multiple copies of integrated HBV DNA. The integrated HBV copy number contained in these cell lines was determined by Gough and Murray, (1982). The cell line Rat2/130.4/TK4 contains approximately ten copies of the HBV genome per haploid rat genome, whereas, the L/130.4/TK154 cells contain approximately one intact HBV genome per haploid genome. The mouse L cell line is partially tetraploid and therefore the number of intact HBV genomes per cell is probably two to three (Gough and Murray, 1982). This estimate of integrated HBV copy number contained in the mouse L cell line is probably an underestimate of the true copy number, since one of the clones isolated, ALTK154-H7 (see below), contains a complete HBV genome. However, the corresponding fragment on the Southern blot of Hind III digested L/130.4/TK154 DNA is not one of the major hybridising bands, indicating that the cells probably contain more than two or three integrated copies of HBV.

Since the pattern of integration of the mouse L cell line L/130.4/TK154 is less complex than the corresponding pattern of Rat2/130.4/TK4, the mouse L cell line was chosen for this study. Both cell lines express the HBsAg and HBcAg/HBeAg at equivalent levels (Gough and Murray, 1982).
Figure 3.2. Detection of HBV sequences in genomic DNA. The samples were separated on 0.7% (w/v) agarose gels and hybridised to genomic length HBV DNA. A) DNA from L/130.4/TK154 (10 μg), digested with Eco RI (track 1), Hind III (track 2) or Eco RI & Hind III (track 3). To provide size markers, bacteriophage λcI857 DNA was digested with Hind III, sizes are given (in kb). B) DNA from Rat 2/130.4/TK4 (10 μg), digested with Eco RI. Molecular size markers used were λcI857 DNA digested with Eco RI & Hind III, sizes are given (in kb). C) Southern blot examination of genomic clones, DNA from L/130.4/TK154 (10 μg), digested with Eco RI & Hind III (track 1); DNA from λLTK154-H7 digested with Eco RI & Hind III (track 2); DNA from λLTK154-R3 digested with Eco RI & Hind III (track 3). Molecular size markers used were λcI857 DNA digested with Hind III, sizes are given (in kb). The L/130.4/TK154 Eco RI & Hind III DNA band co-migrating with the λLTK154-R3 fragment is a doublet, the λLTK154-R3 fragment does not contain a Hind III restriction site.
The HBV subtype cloned in pHBV130 and pHBV130-4 does not contain either Eco RI or Hind III restriction endonuclease cleavage sites. Therefore, these enzymes were used for generation of the genomic libraries. Digestion of the genomic DNA from the cell line L/130-4/TK154 with either, or both, of these enzymes results in a complex pattern of HBV hybridising bands when analysed by Southern blotting (fig. 3.2). The two most strongly hybridising Eco RI fragments are too large to be conveniently cloned into the vector λEMBL4. However, digestion with Hind III or double digestion Hind III and Eco RI and Southern blot analysis indicates that the HBV sequences contained within the two largest Eco RI fragments are contained within the two largest Hind III fragments which are of clonable size. Therefore, genomic libraries were constructed by digestion of the genomic DNA with either Eco RI or Hind III and cloning the resultant DNA fragments into the lambda vectors λEMBL4 and λCh34 respectively.

The lambda vectors λEMBL4 and λCh34 can both accommodate DNA inserts of approximately 9 to 20 kb. Both vectors were constructed with inverted polylinker sequences between the lambda arms and the stuffer fragments (Frischauf et al., 1983; Loenen and Blattner, 1983). By using double restriction endonuclease digestions it is possible to generate lambda arms and stuffer DNA fragments with different ends, thus preventing the stuffer fragments from religating to the vector arms. Digestion of λEMBL4 with Eco RI and Bam HI or λCh34 with Hind III and Bam HI followed by isopropanol precipitation of the nucleic acids, leaves the short cohesive linker fragments in the supernatant, thereby eliminating the possibility of religation of the middle fragment to the lambda arms (Frischauf et al., 1983). Using this procedure it is possible to selectively enrich for recombinants by ligation of DNA fragments containing the same cohesive ends to the lambda arms. The vector λEMBL4 was also constructed such that only recombinants will grow on a P2 lysogen (Frischauf et al., 1983). The middle fragment of λEMBL4 contains the lambda red and gam genes which provides a genetic selection. Growth of the Spf" vector is selected against when plated on a E. coli P2 lysogen. Chimeric constructions, which contain a foreign
insert, are effectively red⁻ and gam⁻, therefore the Spt⁻ recombinants will grow on the P2 lysogen.

3.1.2 Construction of the Genomic Libraries

The L/130•4/TK154 chromosomal DNA used for the library constructions was obtained from laboratory stocks. Fifty micrograms of chromosomal DNA was digested with either Eco RI or Hind III and size fractionated on a 10% to 40% (w/v) sucrose gradient. The fractions were analysed by agarose gel electrophoresis and those containing DNA fragments, from 6 kb to less than 20 kb, were pooled, dialysed against 10 mM Tris-HCl pH 7.5, 1 mM EDTA to remove the sucrose and then ethanol precipitated (data not shown).

The genomic libraries were constructed by ligation ca. 0.5 μg of digested chromosomal DNA and 1.5 μg of doubly digested, heat inactivated (65°C for 20 minutes) and isopropanol precipitated lambda vector DNA in a total volume of 20 μl. The Eco RI generated genomic library using the vector λEMBL4 was tested for the efficiency of generation of recombinants by in vitro packaging the equivalent of ~0.1 μg of vector DNA and plating on E. coli 5K (hsdR) and Q359 (hsdR, P2 lysogen). Only recombinants should plate on Q359, whereas, both vector and recombinants will grow on 5K. One hundred plaques were picked at random from the bacteriophages grown on 5K and replica plated onto separate lawns of 5K and Q359. Routinely, 70 to 80% of the phage grown on strain 5K grew on Q359, indicating that 70 to 80% of the packaged lambda DNA consisted of recombinant molecules. Although the E. coli strain Q359 specifically selects for Spt⁻ recombinants, it was observed that the efficiency of plating in vitro packaged recombinant phage molecules was reduced approximately two fold when plated on Q359, relative to plating on 5K (data not shown). Therefore, both the Eco RI and Hind III generated genomic libraries were in vitro packaged and plated onto E. coli strain 5K.
The Eco RI and Hind III genomic libraries were screened by plaque hybridisation using $^{32}$P-labelled pHBV130·4. After screening approximately $10^6$ recombinant plaques from each library, one recombinant from each library was isolated that hybridised to the probe. The positive clones were purified and lambda DNA prepared. The clones were found to contain restriction fragments corresponding to the third largest HBV Eco RI fragment and seventh largest HBV Hind III fragment of L/130·4/TK154. These clones were designated XLTK154-R3 and XLTK154-H7 respectively. Double digestion of the clones and L/130·4/TK154 DNA with Eco RI and Hind III and analysis by southern blotting (fig. 3.2), indicates that the cloned fragments do correspond to genuine integrations contained in L/130·4/TK154 and that the clones represent separate integrations.

Further attempts to construct genomic libraries using L/130·4/TK154 DNA were not possible due to exhaustion of the laboratory stocks of the chromosomal DNA, and attempts to recover the L/130·4/TK154 cell line from frozen stocks were unsuccessful, therefore it was not possible to isolate chromosomal DNA from this cell line. The cell line Rat2/130·4/TK4 was the only readily available, viable cell line of those constructed by Gough and Murray, (1982). Genomic DNA was prepared from the Rat2/130·4/TK4 cell line and used to construct Eco RI genomic libraries, as described above, using the vector $\lambda$EMBL4. After screening approximately $5 \times 10^6$ recombinants, 5 positive signals (from the same ligation reaction) were observed when probing the genomic library with $^{32}$P-labelled pHBV130·4. Further analysis of these recombinants (XTK4-1 to XTK4-5) indicated that the two isolates XTK4-3 and XTK4-4 contained only pBR322 derived sequences and were not studied any further (data not shown). Southern blot analysis of the remaining clones and Eco RI digested Rat2/130·4/TK4 DNA, using $^{32}$P-labelled gel purified HBV DNA as a probe, showed that the cloned DNA fragments contained in XTK4-1, XTK4-2 and XTK4-5 did not correspond to restriction fragments of Rat2/130·4/TK4 (data not shown). This indicates that these clones arose by a cloning artifact. Further analysis by Southern blotting using
$^{32}$P-labelled λEMBL4 indicated that the central Eco RI fragment of the clones hybridises to lambda DNA (data not shown).

It is evident that the cloning efficiency of the HBV containing fragments corresponding to genomic integrations of the cell lines L/130·4/TK154 and Rat2/130·4/TK4 was very low. Screening of the number of recombinants isolated should have resulted in the identification of a number of HBV positive clones, particularly from the Rat2/130·4/TK4 cell line. It is possible that the genomic libraries were not completely representative due to the preferential ligation of concatemers of small DNA fragments into the lambda arms. Alternatively, the HBV integrations contained on the restriction fragments, when ligated to the lambda vectors, could be unstable when propagated on the E. coli 5K host, due to repeated HBV-pBR322 or cellular flanking sequences and were therefore lost from the genomic libraries. Several cases have been reported where the presence of tandemly-repeated sequences in cloned eukaryotic DNA has led to deletion by recombination during phage propagation on certain E. coli hosts (Arnheim et al., 1979; Fritsch et al., 1980; Lauer et al., 1980).

If the constructions of the genomic libraries were to be repeated, the libraries could be propagated on a strain deficient in both exonuclease I and V, such as, E. coli CES200 (recBC-, sbcB-, hsdR-) (Nader et al., 1985). Nader et al., (1985) have reported that the strain E. coli CES200 can be used for the propagation of λEMBL3 Spt- recombinants, which are unstable when propagated on other E. coli strains. The strain E. coli CES200 was not available when the genomic libraries were constructed and screened for the presence of integrated HBV DNA sequences.

### 3.2 Analysis of the Genomic Clones

The DNA fragments cloned into the lambda vectors were analysed by heteroduplex electron microscopy, using Eco RI digested pHBV130 (fig. 3.1) as a probe, and by restriction mapping/Southern blot hybridisation, to determine the structure of integrated HBV DNA.
The junctions of the viral and cellular sequences were determined by cloning restriction fragments containing the integration sites into M13 based vectors and determining the nucleotide sequence, as described in materials and methods.

3.2.1 ALTK154-H7

The structure of the Hind III fragment cloned into ALTK154-H7 was determined by heteroduplex hybridisation (fig. 3.3). The heteroduplex analysis indicates that the clone contains approximately 114% of HBV sequences, from position ca. 1000 ± 100 through one HBV genome equivalent, to the original oligoG-C tail Pst I cloning site (ca. 1460 ± 50), plus approximately 1.5 kb of the pBR322 vector sequences integrated into mouse DNA sequences. The structure of ALTK154-H7 was confirmed by restriction endonuclease mapping (fig. 3.4).

The vector λCh34 has the capacity to accommodate DNA fragments of 9 to 20 kb. It is evident that the Hind III fragment cloned into ALTK154-H7 is of the order of 5.6 kb, which would result in the clone being below the optimum packaging size. The Hind III digestion of ALTK154-H7 also generates a fragment of approximately 2.25 kb, which corresponds to the size of one of the vector stuffer Hind III fragments. At least two of these fragments were co-ligated with the HBV containing fragment to generate ALTK154-H7. The stuffer fragments obviously still contain the polylinker attached to their ends, which indicates that the Bam HI restriction reaction did not digest to completion.

The HBV Bam HI site at position 1004 was determined to be contained in the clone with the corresponding terminal Hind III/Bam HI fragment being approximately 250 base pairs. The position of the viral-cellular junction was confirmed by gel purifying the H7–Hind III fragment and re-digesting with Bam HI; the resultant restriction fragments were cloned into appropriately digested mp18 and mp19 vectors. The nucleotide sequence of the 277 base pair Hind III/Bam HI fragment was determined,
Figure 3.3 Mapping of the HBV sequences in the clone XLTK154–H7 by heteroduplex analysis. A) Electron micrograph of a heteroduplex between clone XLTK154–H7 and pHBV130 (Eco RI). B) Interpretive drawing of the electron micrograph. Thin and thick lines represent single-stranded and double-stranded DNA, respectively. C) A schematic representation of the heteroduplex structure. Approximate sizes of single-stranded and double-stranded regions are given (in kb). The HBV-cellular junction is indicated by approximate HBV co-ordinates (below).
Figure 3.4 Restriction endonuclease cleavage patterns of λLTK154-H7. Approximately 0.5 μg of DNA was digested with Hind III (track 1), Hind III & Eco RI (track 2), Hind III & Xho I (track 3), Hind III & Bst EII (track 4), Hind III & Bam HI (partial digestion) (track 5), Hind III & Bgl II (track 6) and Hind III & Pst I (partial digestion) (track 7).

A) The DNA fragments were separated on a 0.7% (w/v) agarose gel. Size markers were Hind III digested λcI857 DNA (track M), sizes are given (in kb). B) The DNA fragments probed with a radioactively labelled Xho I HBV DNA fragment of pHBV130. C) The DNA fragments probed with radioactively labelled pBR322 DNA.
indicating that the viral-cellular junction is positioned at nucleotide 994 on the HBV130 map (fig. 3.5).

The λLTK154-H7 Hind III fragment (fig. 3.12) contains the complete coding sequences for the preS2/surface gene, the X ORF and the preC/core gene. Cattaneo et al., (1983b) have postulated that the major surface promoter is contained between HBV nucleotides 1188 and 1260, therefore, the H7 DNA fragment should be capable of directing transcription of the surface gene, with the mRNAs being processed at the polyadenylation signal at position 16 (Gough, 1983). Also contained on the fragment is the complete putative transcriptional unit for the X mRNA (Gough, 1983). Although the fragment contains the HBV enhancer, core promoter and the complete preC/core gene, it is unlikely to direct synthesis of the core protein, due to the lack of an appropriately placed polyadenylation signal (Cattaneo et al., 1984). Synthesis of the core mRNAs would require co-transcription into pBR322 and also possibly mouse sequence, until a suitable polyadenylation signal was transcribed.

3.2.2 λLTK154-R3

The pattern of heteroduplexes obtained for the clone λLTK154-R3 was very complex due to directly repeating HBV and pBR322 sequences contained on the integration. One of the heteroduplexes obtained is shown in fig. 3.6. The pHBV130 molecule hybridised to the λLTK154-R3 molecule shown in fig. 3.6, is deleted by approximately 1.6 kb from the end of the molecule due to fragmentation, possibly resulting from a single strand break. The presence of the ~0.5 to 1.0 kb single strand end of the Eco RI digested pHBV130 probe, indicates that the cloned sequences contain non-homologous sequences to this region, resulting from either deletion or inversion of the pBR322 sequences (see below). One of the two possible orientations of hybridisation is shown in fig. 3.6. The structure shown was confirmed by restriction mapping and Southern blot hybridisation.
a) \textit{XLTK154-H7 Junction Sequence}

\begin{verbatim}
Hind III . . . . Eco RI .
AAGCTTATTATATGGCAGATTCTTTACTGAGAT
 . . . . Eco RI.
ACTTAGGCTCGTTTCTTAAATGTGTGTACATATTGAAAGCACACTCA
 . . . . Eco RI.
TACAAAGCATGTCCCATGGCAGATGCACTTGAGATTCGGCTAGAAATTTTGAT
TCCATTGCTGAAATGTTTTCTATATCCGACAGTCACCCTATTACTAGC
 . . . . Eco RI.
GCTGCCTCATTGGGAATGAACGTAAGCATGTCCCATTGGAAGTCACTGAATTCGCCTAGAATTTTCGAT
TCCATTCGTGAAAGGTGGGTTCTATATCCCGACAGTCACCCTATTACTAGC
mouse 994 HBV Bam HI .
CACAACAAATTTACCACCGAGTTGATGCAGCCTCTGCAGGACAAACACAAACA
\end{verbatim}

b) \textit{XLTK154-R3 Junction Sequence}

\begin{verbatim}
1409 Xho I . . . . .
CTCGAGGTTGGGACCCTGCGCTGAACATGGGAGAACATCACATCAATT
 . . . . . .
cctaggacccctgtcggttatccaggggcgggtttttccgttgacaaaga
 . . . . . .
tccctcacaattaccgcagagtctagactcggtggagtctctgctcattttt
 . . . . . .
ctagggggaactaccggtgtgtcttggccaaatctggcatctcacaattctc
 . . . . . .
caatcactcaccaacctctgtctctcaaccttcgctgtgtaatctgcggga
HBV 1655 mouse . Xho I.
TGTGTCAGGGCACGAAAAATGAAGTGCGGCTGTCGCGTGCGG
\end{verbatim}

\textbf{figure 3.5} Nucleotide sequence (HBV plus strand) of the HBV-cellular junctions of a) XLTK154-H7 and b) XLTK154-R3. Shown are relevant restriction endonuclease sites. Integration sites are indicated by a vertical bar above the sequence corresponding to the terminal HBV nucleotide.
Figure 3.6 Mapping of the HBV sequences in the clone λLT154-R3 by heteroduplex analysis. A) Electron micrograph of a heteroduplex between clone λLT154-R3 and pHBV130 (Eco RI). B) Interpretive drawing of the electron micrograph. Thin and thick lines represent single-stranded and double-stranded DNA, respectively. C) A schematic representation of the heteroduplex structure. Approximate sizes of single-stranded and double-stranded regions are given (in kb). The discontinuities with the pHBV130 probe are indicated by approximate HBV co-ordinates (below).
The restriction mapping and blot hybridisation of the clone XLTK154-R3 (fig. 3.7) was more informative than the heteroduplex analysis. The presence of the 0.75 kb Eco RI/Pst I fragment, which specifically hybridises to $^{32}$P-labelled pBR322 DNA, corresponds to the size of the pBR322 restriction fragment (pBR322 co-ordinates 3608-4361), indicating that the Eco RI cloning site is derived from integrated pBR322 sequences. The 0.56 kb Pst I fragment, which specifically hybridises to $^{32}$P-labelled HBV DNA, corresponds to the expected size of the HBV insert present between the two pBR322 vector fragments of pHBV130.4. The presence of the 1.3 kb (Eco RI/Xho I) and 0.85 kb (Eco RI/Bam HI) fragments, which hybridise to both $^{32}$P-labelled pBR322 and HBV DNAs, is consistent with the pBR322 Eco RI cloning site. The failure of the 0.75 kb (Eco RI/Bst EII) fragment to hybridise to HBV DNA is consistent with the original oligoG·C tailing cloning site of pHBV130, being immediately upstream of the Bst EII (915). Using a higher specific activity $^{32}$P-labelled HBV probe, this fragment was shown to weakly hybridise to HBV sequences (data not shown). The internal Pst I fragment of 3.8 kb, which hybridised strongly to pBR322 DNA, is smaller than the expected length of pBR322 (4.3 kb), indicating that the pBR322 sequences contain a ca. 0.5 kb deletion. The presence of the 4.4 kb fragment in the digestions with Xho I, Bst EII and Bam HI, which hybridises to both probes, is also consistent with the pBR322 sequences containing a ca. 0.5 kb deletion.

If the Pst I fragments of this clone were representative of fragments of pHBV130.4, it would be expected that the fragments should specifically hybridise to either HBV or pBR322 DNA. The 3.8 kb and 1.95 kb Pst I fragments, shown in fig. 3.7, hybridise with both HBV and pBR322 DNA. Hybridisation of HBV DNA to the 3.8 kb fragment is not an artifact due to vector contamination of the HBV probe, since strand specific $^{32}$P-labelled HBV DNA (the Xho I genomic length HBV fragment cloned in M13mp18) also weakly hybridised to this band (data not shown). The hybridisation pattern is consistent with a small inversion of the pBR322 and HBV sequences surrounding the Pst I site. Since the overlapping 4.4 kb Bst EII, Bam HI and Xho I fragments are all of the same size, this indicates that the inversion occurred between the Bst EII and Pst I
Figure 3.7 Restriction endonuclease cleavage patterns of λLTK154-R3. Approximately 0.5 μg of DNA was digested with Eco RI (track 1), Eco RI & Hind III (track 2), Eco RI & Xho I (track 3), Eco RI & Bst EII (track 4), Eco RI & Bam HI (partial digestion) (track 5), Eco RI & Bgl II (track 6), Eco RI & Pst I (track 7). A) The DNA fragments were separated on a 0.7% (w/v) agarose gel. Size markers were Hind III digested λcI857 DNA (track M), sizes are given (in kb). B) The DNA fragments probed with a radioactively labelled Xho I HBV DNA fragment of pHBV130. C) The DNA fragments probed with radioactively labelled pBR322 DNA.
sites. It should be noted that the Eco RI, Hind III and Bam HI sites are deleted from the internal pBR322 sequences. The non-homologous region observed in the heteroduplex analysis probably corresponds to the inverted sequences, or both the inverted and deleted sequences.

The Xho I fragment (≈ 0.3 kb) which hybridises to HBV DNA, does not correspond to a restriction fragment of pHBV130-4, therefore, a viral-cellular junction probably occurs within this fragment. The 1.95 kb Pst I fragment was gel purified and digested with Xho I. The resultant fragments were ligated to Sal I digested mp18 DNA and the nucleotide sequence of the 294 base pair Xho I fragment determined (fig 3.5). The nucleotide sequence of the small Xho I fragment indicates that the viral-cellular junction occurs at HBV position 1655. The restriction map of λLTK154-R3 is shown in fig. 3.12.

3.2.3 λTK4-1 and λTK4-2

As indicated above, the Rat2/130-4/TK4 genomic clones isolated are the result of a cloning artifact. No heteroduplexes were obtained for either of the clones λTK4-1 or λTK4-2, possibly due to the limited sequence homology between these clones and pHBV130. The structure of these two clones was determined by restriction mapping and Southern blotting (fig. 3.8, λTK4-1; fig. 3.9, λTK4-2). As with λLTK154-R3, the Eco RI cloning site of λTK4-1/2 corresponds to the pBR322 Eco RI site, with the cloned sequences being discontinuous with HBV downstream of the Bam HI (1004) site (fig. 3.12). The 2.2 kb Bst EII fragments from both clones, which hybridise to 32P-labelled HBV DNA, were gel purified. The ends of the fragments were "filled in" using the Klenow fragment DNA polymerase and ligated to Sma I digested mp18. The nucleotide sequences of the Bst EII fragments are consistent with the HBV DNA being digested with Bam HI and ligated to the Bam HI site of the polylinker sequence of λEMBL4; with the Bam HI site being immediately followed by a Sal I site and vector sequences (data not shown). It is evident that the ligation reaction used to generate the genomic library contained a residual Bam HI activity that was ineffectively inactivated by heat treatment.
Figure 3.8 Restriction endonuclease patterns of λTK4-1. Approximately 0.5 μg of DNA was digested with Eco RI (track 1), Eco RI & Hind III (track 2), Eco RI & Xho I (track 3), Eco RI & Bst EII (track 4), Eco RI & Bam HI (track 5), Eco RI & Bgl II (track 6), Eco RI & Pst I (track 7).

A) The DNA fragments were separated on a 0.7% (w/v) agarose gel. Size markers were Hind III digested λcI857 DNA (track M), sizes are given (in kb). B) The DNA fragments probed with a radioactively labelled Xho I HBV DNA fragment of pHBV130. C) The DNA fragments probed with radioactively labelled pBR322 DNA.
Figure 3.9  Restriction endonuclease patterns of λTK4-2. Approximately 0.5 μg of DNA was digested with Eco RI (track 1), Eco RI & Hind III (track 2), Eco RI & Xho I (track 3), Eco RI & Bst EII (track 4), Eco RI & Bam HI (track 5), Eco RI & Bgl II (track 6), Eco RI & Pst I (track 7). A) The DNA fragments were separated on a 0.7% (w/v) agarose gel. Size markers used were Hind III digested λcI857 DNA (track M), sizes are given (in kb). B) The DNA fragments probed with a radioactively labelled Xho I HBV DNA fragment of pHBV130. C) The DNA fragments probed with radioactively labelled pBR322 DNA.
3.2.4 λTK4-5

The clone λTK4-5 was analysed by heteroduplex analysis (fig. 3.10). The heteroduplex analysis indicates that the clone contains approximately 1.9 kb of homology with pHBV130. One possible heteroduplex orientation is shown in fig. 3.10, which indicates that the cloned sequences show discontinuities with the HBV sequences at positions 1000 ± 100 and 2300 ± 100. The structure was confirmed by restriction mapping and Southern blot hybridisation (fig 3.11). The 32P-labelled pBR322 probe failed to hybridise to the restriction fragments of this clone, indicating that the cloned fragment does not contain pBR322 sequences (data not shown). The restriction map of λTK4-5 is shown in fig. 3.12.

The 2.2 kb Bst EII fragment from λTK4-5 was gel purified, the ends "filled in", cloned into Sma I digested mp18 and the nucleotide sequence determined. Like the clones λTK4-1 and λTK4-2, the sequences downstream of the HBV Bam HI (1004) site of λTK4-5, correspond to vector sequences (data not shown). The integration site, at the left end of the clone, is not represented by a Bam HI site and therefore, probably is a true integration site. Attempts to determine the sequence of the integration site were unsuccessful. The integration site is contained on the 1.0 kb Eco RI/Bam HI (HBV position 2682) fragment. This fragment was isolated and digested with Stu I (2251, 2396) and the resultant fragments cloned into Sma I/Bam HI digested mp19, Sma I/Eco RI and Sma I digested mp18. The ligated molecules were used to transform E. coli NM522 and the resultant plaques screened by plaque hybridisation to 32P-labelled HBV DNA. The plaques giving positive signals upon hybridisation were sequenced. Only fragments corresponding to HBV sequences 2396 to 2682 were identified and sequenced. Repeating the sub-cloning obtained the same result. No recombinants corresponding to internal Stu I or terminal Eco RI/Stu I fragments were identified. This could possibly indicate that the HBV-cellular junction occurred immediately upstream of the Stu I (2396) site, with the number of nucleotides being homologous to the probe being limited, reducing the efficiency of hybridisation. The results of the heteroduplex analysis
Figure 3.10 Mapping of the HBV sequences in the clone λTK4-5 by heteroduplex analysis. A) Electron micrograph of a heteroduplex between clone λTK4-5 and pHBV130 (Eco RI). B) Interpretive drawing of the electron micrograph. Thin and thick lines represent single-stranded and double-stranded DNA, respectively. C) A schematic representation of the heteroduplex structure. Approximate sizes of single-stranded and double-stranded regions are given (in kb). The discontinuities with the pHBV130 probe are indicated by approximate HBV co-ordinates (below).
A

B

C

pHBV130

\[ \text{\sim} 2.2 \text{ kb} \quad \text{\sim} 4.1 \text{ kb} \]

\[ 2300 \pm 100 \quad 1000 \pm 100 \]

(HBV) (HBV)
Figure 3.11 Restriction endonuclease patterns of λTK4-5. Approximately 0.5 μg of DNA was digested with Eco RI (track 1), Eco RI & Hind III (track 2), Eco RI & Xho I (track 3), Eco RI & Bst EII (track 4), Eco RI & Bam HI (partial digestion) (track 5), Eco RI & Bgl II (track 6), Eco RI & Pst I (track 7). A) The DNA fragments were separated on a 0.7% (w/v) agarose gel. Size markers used were Hind III digested λcI857 DNA (track M), sizes are given (in kb). B) The DNA fragments probed with a radioactively labelled Xho I HBV DNA fragment of pHBV130.
Figure 3.12  Restriction endonuclease maps of the genomic clones isolated from the cell lines L/130·4/TK154 and Rat2/130·4/TK4.

The shaded blocks represent HBV sequences, the open blocks represent pBR322 sequences, solid thin lines represent eukaryotic sequences and dashed horizontal lines represent lambda sequences.  a) λLTK154-H7, showing the HBV-cellular integration site at position 994.  Shown above the map are the HBV transcriptional regulatory signals: SPI, preS1 gene promoter; SPII, preS2/surface gene promoter; En/XP, HBV Enhancer region/X gene promoter region; CP, core promoter region; poly(A), polyadenylation signal.  Shown below the map are the relative positions of the viral open reading frames: C, core gene; P, polymerase gene; preS/S, preS/surface gene, X, X gene.  b) λLTK154-R3, showing the HBV-cellular integration site at position 1655.  Also shown is the HBV-pBR322 inversion surrounding the Pst I site, above the map is shown the approximate position of the ca. 0·5 kb pBR322 deletion.  c) λTK4-1, showing the position of the ligation of the HBV sequences (position 1004, Bam HI) to bacteriophage lambda (λ) sequences (dashed line).  d) λTK4-2, showing the position of the ligation of HBV sequences (position 1004, Bam HI) to bacteriophage λ sequences (dashed line).  e) λTK4-5, showing the approximate position of the HBV-cellular integration at HBV position 2300 ± 100.  Also shown is the position of ligation of HBV sequences (position 1004, Bam HI) to bacteriophage λ sequences (dashed line).  Shown below the map is the relative positions of the viral open reading frames contained on this clone (see fig. 3.12a).  Restriction endonuclease abbreviations: B, Bam HI; Bg, Bgl II; Bs, Bst EII; E, Eco RI; H, Hind III; Ps, Pst I and Xh, Xho I.
are consistent with this hypothesis, suggesting that the integration site is at approximate HBV position 2300 ± 100.

3.3 Discussion: Integrated HBV DNA in Hepatocellular Carcinomas and Derived Cell Lines

A model of HBV integration has been proposed by Koshy et al., (1983) and Koch et al., (1984b). These authors observed that three HBV integrations isolated from the PLC/PRF/5 cell line contained the HBV-cellular integration events within the single stranded region present in the Dane particle HBV genome. Their hypothesis assumes that replicative intermediates within the nucleus of the infected cell contain regions of single stranded DNA analogous to the Dane particle DNA. The model proposes that the integrative process interacts with cellular DNA replication. Interaction between the HBV genome and cellular sequences could possibly be facilitated by transient alignment of short stretches of nucleotide homology between the HBV and cellular sequences (Koch et al., 1984b). DNA polymerases using the cellular template are proposed to switch to the single stranded gap region of the HBV genome close to the 5' end of the short strand. A recombination event then joins the long strand with the cellular sequences, which was previously the template for DNA polymerase. Finally, the single-stranded regions are filled and ligated to yield an integrated HBV DNA molecule. The authors propose it is possible that this mechanism also uses other single-stranded regions of the viral genome; either the naturally occurring "nick" or randomly generated nicks. If this model is correct, HBV integration would take place preferentially via the large single strand gap.

A more extensive analysis of integrated HBV DNA present in the Alexander cell line (Shaul et al., 1984; Ziemer et al., 1985), indicates that the viral-cellular junctions are essentially dispersed over the entire HBV genome, with some preference for integration to occur within the double-stranded region of the genome. Therefore, these results are inconsistent with the above model. The results obtained in this study
are consistent with this observation of random integration with no specific viral sequences required for integration; with integration sites at HBV nucleotides 994, 1655 and 2300 ± 100. However, it is probably not a reasonable assumption that the integration pattern observed in this study represents natural integration events, due to structural differences between the transforming DNA used to construct L/130·4/TK154 and Rat2/130·4/TK4 and the native HBV genome. Gough and Murray (1982) used pHBV130·4, which consists of four fully double-stranded head-to-tail copies of HBV DNA cloned between two copies of pBR322, to transform mouse L and Rat2 cells using the calcium-phosphate procedure. The naturally occurring intracellular HBV DNA exists in two forms: i) the nuclear monomeric covalently closed double stranded DNA which is used as transcriptional templates, and ii) the cytoplasmic replicative forms with structures reminiscent of the Dane particle DNA. The observations of Dejean et al., (1984), Hino et al., (1986) and Yaginuma et al., (1987a), indicate that semi-specific integration can occur via the DR sequences, which probably represents a recombination event via the free ends of replicative intermediates. The DNA molecules used by Gough and Murray (1982) did not contain these free ends, therefore recombination in this manner would not have occurred.

The strongest impression obtained from the several detailed descriptions of integrated DNA from primary tumours (Dejean et al., 1984, 1986; Rogler et al., 1985; Fowler et al., 1986; Hino et al., 1986; Yaginuma et al., 1987a) and hepatoma derived cell lines (Dejean et al., 1983; Kolke et al., 1983; Koshy et al., 1983; Koch et al., 1984b; Shaul et al., 1984, 1986b; Mizusawa, et al., 1985; Yaginuma et al., 1985; Zeimer, 1985; Berger and Shaul, 1987), is one of irregularity. The integrated viral DNA may be co-linear with the genome or rearranged with one or more deletions or inversions. The cellular sequences that serve as integration sites show no obvious relationship to each other, nor to the recombination sites within the viral DNA. The host integration sites may be located on any of several chromosomes, with the chromosomes exhibiting deletions, inversion, duplication and trans-locations. All of these features are reminiscent of nonhomologous events that mediate integration of experimentally transfected DNA into cellular chromosomes. Therefore, the
study of experimentally integrated DNA may exemplify the integration mechanisms of HBV DNA.

The microinjection studies of Folger et al., (1982) indicate that linear DNA molecules have an integration efficiency 40-fold greater than circular molecules. Examination of integrated sequences indicates that recombination events occur at random when using circular molecules. However, linear molecules tend to integrate at the ends of the fragments with very little loss of terminal sequences, however, the original restriction endonuclease sites are deleted. These results suggest that linear molecules are the preferred substrate for illegitimate recombination of exogenous DNA in the mammalian genome and that such molecules are integrated via their ends. In support of this model, it has been shown that mammalian cells can introduce double-strand breaks in exogenous DNA (Wake et al., 1984) and that free ends of DNA are efficiently ligated (Folger et al., 1982; Wilson et al., 1982; Wake et al., 1984). It has been proposed that the principal mechanism of nonhomologous recombination in transfected DNA is end-to-end ligation (Wilson et al., 1982; Wake et al., 1984). Wake et al., (1984) propose that integration of foreign DNA into host chromosomes, may possibly occur by ligation of the foreign DNA into transient chromosomal breaks.

Bullock et al., (1985) have proposed that eukaryotic topoisomerase type I may be involved in illegitimate recombination. Topoisomerase I reactions involve double-stranded DNA substrates which are nicked at discrete loci, with the enzyme remaining bound to the DNA by a covalent linkage to the 3'-phosphate at the site of the break. The closing of the nick is mediated by topoisomerase I by ligation of the 3'-phosphate to the 5'-hydroxyl. Bullock et al., (1985) propose that the religation reaction may not always be proficient and that free 3' ends containing bound topoisomerase I could be capable of ligation to other DNA strands containing a free 5'-hydroxyl. Nicking by topoisomerase I of DNA, near a pre-existing nick or gap on the opposite strand, could generate a double strand break if the number of complementary nucleotides between the strand breaks were insufficient to hold the double stranded structure together. Ligation utilising the topoisomerase I activity of the double
strand breaks to chromosomal nicks, possibly also generated by topoisomerase I, could facilitate illegitimate recombination. The ligation of exogenous DNA into the chromosomal DNA need only account for one strand of the junction sequence. The single strand could give rise to complementary sequences by serving as a template for subsequent rounds of DNA replication or DNA repair. This model of Bullock et al., (1985) is essentially consistent with the proposed models of Folger et al., (1982) and Wake et al., (1984).

If the above models do represent the mechanism of HBV DNA integration, it is conceivable that random double stranded breaks are introduced into the nuclear covalently closed HBV DNA, which then recombine, by ligation mechanisms, with cellular sequences via transient breaks in the genomic DNA. The observation that HBV DNA integrates via the DR sequences is also consistent with the above hypothesis. Tuttleman et al., (1986) have shown that the nuclear supercoiled HBV DNA is generated by de novo synthesis in the cytoplasm and not by semi-conservative replication. Therefore, it is possible that open circular replicative forms are transiently present in the nucleus. The open circular molecules would need to be processed to remove the protein and RNA primers and the terminal redundancy [r] before ligation of the viral strands. Therefore, replicative intermediates will contain free DNA ends which could integrate via the above proposed model. It is possible that single strand breaks, in either the gap region or opposite the "nick" of replicative intermediates, would also generate linear substrates for integration which is consistent with the observation of integration occurring within the cohesive end region (Nagaya et al., 1987).

Yaginuma et al., (1987a) have observed that one isolated HBV integration from a tumour appears to have integrated via the terminal redundancy present on the 5' and 3' ends of the minus strand of a replicative intermediate. The integration event resulted in the loss of two-three nucleotides from the 3' end of the minus strand, with a concomitant loss of 16 base-pairs from the chromosomal DNA at the site of integration. Dejean et al., (1984) have suggested that homologous recombination events may be involved in HBV integration, with a specific recombination
event occurring within the direct repeat sequences and chromosomal DNA. The chromosomal clone of Yaginuma et al., (1987a), corresponding to the integration site of the above HBV integration, showed no significant sequence homology between viral and chromosomal DNA sequences, therefore, homologous recombination is probably not a general mechanism involved in HBV integration.

If semi-specific integration does occur via the free ends of replicative intermediates, the DR2 integration event observed by Dejean et al., (1984,1986) would require transient formation of a covalent RNA-DNA hybrid to produce the observed junction, since the site of integration corresponds to the putative RNA primer sequence. Therefore, it is possible that this integration site does not represent an event mediated by the free DNA end of a replicative intermediate, but that it could represent, a random integration site which is fortuitously placed within the DR2 sequence.

The majority of the observed patterns of HBV integration are consistent with random integration events. Therefore, it is possible that the initial stages of integration may involve the generation of random, single or double strand breaks into the HBV genomic DNA, which then recombine with the cellular genome via single strand ligations to genomic "nicks", rather than template switches during chromosomal replication as proposed by Koshy et al., (1983) and Kock et al., (1984b).
Chapter Four
Results and Discussion

Transcriptional Control of the Major Surface Promoter

4.1 Analysis of HBsAg Expression Directed by Chromosomally Integrated HBV DNA

4.1.1 Transfection of Eukaryotic Cells with XLTK154-H7

Analysis of the genomic clones isolated in bacteriophage lambda vectors, indicated that only the clone XLTK154-H7 contains a complete HBV integration capable of expressing native HBV transcripts. The viral/cell DNA junction at position 994 on the HBV map is situated between the preS1 "TATA" box promoter (position 876-882) (Rall et al., 1983) and the putative preS2/surface "SV40-late-like" promoter (positions 1188-1260) (Cattaneo et al., 1983b). Therefore, if the results of Standring et al., (1984) are correct, in that the preS2/surface promoter can act independently of the preS1 promoter, the clone XLTK154/H7 should be capable of directing transcription of the major surface mRNA from the preS2/surface promoter. To confirm the results of Standring et al., (1984), the clone XLTK154-H7 was used to transfect the HPRT deficient cell line RJK88 by co-transfection of XLTK154-H7 DNA with the mouse HPRT mini-gene contained on the plasmid pDWM1 (Melton et al., 1986). The mouse HPRT mini-gene consists of a chimeric HPRT gene, constructed by ligation of restriction fragments from genomic and cDNA clones to remove most of the introns. The 33 kb mouse HPRT gene was reduced in size to a 3 kb mini-gene containing the mouse HPRT promoter, the entire 5' and 3' untranslated regions and the coding sequence interrupted by the last two introns (Melton et al., 1986).

Calcium phosphate transformations were conducted in duplicate as described in materials and methods; 5 μg of XLTK154-H7 DNA was digested with Hind III and mixed with 50 ng of Eco RI digested pDWM1 (ie. at an
approximate ratio of the linear Hind III H7 fragment (~ 5.6 kb) to Eco RI digested pDWMI (~ 5.9 kb) of 10:1 molar ratio). As a positive control for transformation efficiency 50 ng of Eco RI digested pDWMI was mixed with 5 μg of sonicated calf thymus DNA (sCTD). Both DNA preparations were used to transform the cell line RJK88. The transformed cells were selected by growth in DMEM/HAT medium. The duplicate XLTK154-H7 co-transformations yielded 4 and 3 HAT resistant colonies each (60 to 80 HAT resistant colonies/μg of pDWMI); the cell cultures were designated RJK88/H7-1 and RJK88/H7-2 respectively. The control pDWMI/sCTD transformation yielded 31 HAT resistant colonies (620 HAT resistant colonies/μg of pDWMI); a negative control culture from this transformation was maintained and designated RJK88/HPRT.

It is evident that the presence of the Hind III digested XLTK154-H7 DNA in the transformation mixture reduced the efficiency of transformation. It was observed that during the transformation procedure, the samples containing XLTK154-H7 DNA formed large aggregated calcium phosphate precipitates on the surface of the cells. Whereas, with the samples containing pDWMI/sCTD the precipitates were smaller and evenly dispersed over the surface of the cells. It is possible that the presence of the bacteriophage λ arms of 20 and 9 kb (approximately 29 kb when joined by annealing of the cos ends) being large DNA molecules could cause aggregation of the calcium phosphate precipitates preventing even dispersion of the precipitate, and therefore, reducing the transformation efficiency by limiting the number of cells exposed to the transforming DNA.

4.1.2 Analysis of Integrated DNA in the Transformed Cells

Total chromosomal DNA was prepared from the RJK88/H7-1 and RJK88/H7-2 cultures; 10 μg of DNA was digested with Bam HI and the DNA fragments separated by agarose gel electrophoresis. The DNA was transferred to nitrocellulose filters and probed with 32P-labelled Xho I digested genomic length HBV DNA. After one weeks exposure, very faint signals corresponding to HBV Bam HI fragments of 1.5 and 1.7 kb were visible on
the autoradiograph, indicating that cell line RJK88/H7-1 contained integrated HBV DNA. DNA from the cell line RJK88/H7-2 showed no hybridisation to the probe (data not shown). The southern blot was repeated using 100 μg of RJK88/H7-1 DNA per lane; the DNA was digested with restriction enzymes that would give diagnostic restriction patterns for the H7 DNA fragment (fig. 4.1a). The southern blot shows that digestion of RJK88/H7-1 DNA with the restriction enzymes generated the corresponding major HBV hybridising fragments of 3·2 kb (Xho I), 1·5 and 1·7 kb (Bam HI) and 2·8 kb (Pst I). All these fragments correspond to internal restriction fragments of the H7 clone indicating that at least one complete copy of the HBV H7 sequences between positions 1004-3182/1-(1460 ± 50) is integrated into the genome of the cell line. The presence of the 5·6 kb Hind III fragment, which hybridises to HBV DNA is suggestive that an uncut XLTK154-H7 molecule may have integrated into the genome of the cells. Another possibility is that the Hind III digested fragments formed concatemers by end-to-end ligation prior to the integration event (Wilson et al., 1982; Folger et al., 1982), thus preserving the Hind III restriction sites.

4.1.3 Analysis of HBV mRNA Directed by the Integrated HBV Sequences

Poly(A)+ RNA was prepared at the same time as the genomic DNA. The poly(A)+ RNA, from the equivalent of 10 Petri dishes (10 cm diameter), was separated on a 1·3% agarose/formaldehyde gel. Poly(A)+ RNA from one dish of the cell line Rat2/130·4/TK4 was separated on the same gel as a positive control. The RNA was transferred to a Hybond-N membrane and probed with 32P-labelled genomic length HBV DNA (fig. 4.1b). The autoradiograph shows that the cell line Rat2/130·4/TK4 produces a HBV mRNA of the expected size for the HBsAg mRNA (2·3 kb; fig. 4.1b, lane 2) (Gough, 1983). In the adjacent lane (RJK88/H7-1; fig. 4.1b, lane 3) is a very faint signal corresponding to a mRNA migrating slightly slower than the Rat2/130·4/TK4 surface mRNA. It is possible that this mRNA also corresponds to the surface mRNA. The slower rate of migration may be due to an artifact of the electrophoresis system since at least 10-fold as much RNA was loaded onto this lane, with the large quantity of RNA
Figure 4.1 Examination of HBV nucleic acids contained in the cell line RJK88/H7-1.

a) Southern blot of RJK88/H7-1 chromosomal DNA (100 µg/track) digested with *Hind* III (track 1), *Xho* I (track 2), *Bam* HI (track 3) and *Pst* I (track 4). The DNA digestions were separated on a 0.7% (w/v) agarose gel, transferred to nitrocellulose and probed with a $^{32}$P-labelled *Xho* I digested unit length HBV fragment from pHBV130. Size markers used were *Eco* RI & *Hind* III digested λcI857 DNA, sizes are given (in kb).

b) Northern blot of poly(A)$^{+}$ RNA isolated from tissue culture cells containing integrated HBV DNA. Poly(A)$^{+}$ RNA from the equivalent of a confluent culture of one × 10 cm diameter Petri dish, Rat2 cells (track 1), Rat2/130·4/TK4 cells (track 2); from ten × 10 cm diameter Petri dishes, RJK88/H7-1 (track 3), RJK88/HPRT (track 4). The RNA was separated on a 1.3% (w/v) agarose/formaldehyde gel, transferred to Hybond-N membrane (Amersham) and probed with a $^{32}$P-labelled *Xho* I digested unit length HBV fragment from pHBV130. The position of the arrow indicates the position of a faintly hybridising signal in track 3. Size markers used were mammalian 28 and 18S rRNA (ca. 4700 and 1800 nucleotides).
inhibiting migration. Two other possibilities may account for the slower rate of migration: i) the mRNA could be a co-transcript of hamster/mouse sequences and HBV sequences and is therefore, slightly larger than the native HBV surface mRNA or ii) that the signal corresponds to a local area of high background and does not represent hybridisation to HBV mRNA. The result obtained from the northern blot is suggestive that the SPII promoter is transcriptionally active if taken in conjunction with the radioimmunoassays (RIA) for the presence of HBV surface antigen in the medium (see below).

4.1.4 HBsAg Expression Directed by Integrated HBV Sequences

The medium from the cells used to isolate DNA and RNA was pooled; 200 µl samples were tested in duplicate for the presence of secreted HBsAg by RIA using the Abbott AUSRIA-II assay for HBsAg. A positive to negative signal ratio (P/N) of 2.1, or greater, is considered statistically positive for the presence of HBsAg. The signals obtained for the radioimmunoassay conducted using neat culture medium indicated that the cell line RJK88/H7-1 was not producing significant amounts of HBsAg. The figures obtained for the P/N ratios were below 2.1 and were of the same order as the signals obtained for the negative control RJK88/HPRT (Table 4.1). The cell line RJK88/H7-1 definitely contains integrated HBV DNA and the results of the northern blot were suggestive that the surface gene was being transcribed. It is possible that the level of secreted HBsAg in the medium was below the level of detection of the radioimmunoassay.

The volume of culture medium obtained from the DNA and RNA preparations was in the order of 300 ml. Since the HBsAg is usually present as 22 nm particles, it was reasoned that it may be possible to recover the HBsAg particles from the medium by high speed centrifugation. Two hundred millilitres of the pooled medium were clarified by centrifugation at 5,000 rpm for 10 minutes in a Sorvall GSA rotor to remove cellular debris. The clarified medium was subjected to high speed centrifugation for 12 hours at 23,000 rpm in a MSE 10×100 ml rotor and the pellets
Table 4.1 Detection of HBsAg by Radioimmunoassay

<table>
<thead>
<tr>
<th>P/N</th>
<th>P/N conc. of medium</th>
<th>Total protein conc. of medium mg/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>neat</td>
<td>medium</td>
<td></td>
</tr>
<tr>
<td>RJK88/H7-1</td>
<td>1.2</td>
<td>2.3, 2.4</td>
</tr>
<tr>
<td>RJK88/HPRT</td>
<td>1.2</td>
<td>1.4, 1.4</td>
</tr>
</tbody>
</table>

Radioimmunoassay results for HBsAg present in the culture medium of RJK88 cells containing chromosomally integrated HBV DNA. Duplicate assays were conducted with 200 μl sample volumes using the AUSRIA II-125 (Abbott) HBsAg radioimmunoassay, the figures given are the average of the duplicate assays. (×200) samples were prepared by ultracentrifugation of the culture medium samples and resuspending the pellets in 1/200 of the initial volume of phosphate buffered saline. Total protein concentrations of the (×200) concentrated samples were determined by the method of Lowry et al., (1951).
resuspended in 1 ml of phosphate buffered saline (ie 1/200 of the initial volume) at 4°C for 24 hours. It was not possible to resuspend all the pellet due to the presence of some insoluble material. The concentrated samples were clarified by centrifugation for 15 minutes in a microcentrifuge to remove the insoluble material, and the supernatants transferred to a fresh 1.5 ml microcentrifuge tube. The protein concentration of the resuspended pellets was determined by the Lowry assay as described in materials and methods. The samples were then tested by RIA for the presence of HBsAg, in duplicate, on two separate occasions. The results of these assays are presented in table 4.1. The concentrated sample from the cell line RJK88/H7-1 consistently gave a P/N ratio in the order of 2.3-2.4 which is just above the level of statistical significance for the presence of HBsAg. The level of P/N ratio obtained for RJK88/H7-1 is probably not an artifact due to high levels of protein in the solutions since, the protein concentration of the negative control RJK88/HPRT and RJK88/H7-1 were not significantly different. This most probably indicates that the H7 transformed cell line does express HBsAg at very low levels.

In an attempt to confirm that the signals obtained from the RIAs were indicative of the presence of HBsAg, analysis of the HBsAg proteins was attempted by SDS polyacrylamide gel electrophoresis (SDS-PAGE) and western blotting. Aliquots of the above concentrated protein samples (50 μl) were immunoprecipitated using the double antibody precipitation procedure, as described in materials and methods, using a high titre chimpanzee anti-HBs serum, raised against yeast derived HBsAg (Murray et al., 1984). The immunoprecipitates and 50 μl of non-immunoprecipitated samples were separated by SDS polyacrylamide gel electrophoresis and transferred to nitrocellulose filters. The filters were probed with rabbit anti-HBs, raised against SDS denatured yeast derived HBsAg (P. Wingfield, personal communication), and then incubated with 125I-labelled protein A and exposed to X-ray film. No signals were obtained corresponding to the expected migration position of the major HBsAg proteins, p25 and gp28 (data not shown). The concentrated protein samples (see above) from RJK88/HPRT and RJK88/H7-1 were labelled in vitro with 125Iodine using Iodogen, with approximately 85% incorporation.
The $^{125}$I-labelled proteins were immunoprecipitated as described above and separated by SDS-PAGE. The polyacrylamide gel was dried and exposed to X-ray film; again no HBsAg proteins were observed. As a final attempt to identify the HBsAg proteins, the cellular proteins were labelled in vivo in the presence of L-$^{35}$S]Methionine. Concentrated HBsAg samples were prepared from the medium by ultra-centrifugation, immunoprecipitated and separated by SDS-PAGE as described above; again no labelled HBsAg proteins were observed. The inability to detect the HBsAg proteins using the methods described above is most likely due to the very low abundance of HBsAg in the preparations, being below the detection limits of the procedures.

Due to the very low expression levels of HBsAg the cell line RJK88/H7-1 was not studied any further. It is evident from the Southern blot experiment that the cell line RJK88/H7-1 is a heterogeneous culture with only a minority of cells containing integrated HBV DNA. Attempts to clonally purify the cells containing the HBV integration(s) were unsuccessful. Attempts to repeat the transformation procedure to obtain further cell lines containing the integrated H7 fragment were also unsuccessful. Since the weak signals obtained from the northern blot analysis and HBsAg RIAs could be artifacts of the system, it is not possible to conclusively that the SPII promoter is transcriptionally active in the system studied here. However, the overall results obtained for HBsAg expression by the cell line RJK88/H7-1 are suggestive that the cloned Hind III fragment contained in λLT154-H7 is capable of directing transcription/expression of the HBsAg, indicating that the SPII promoter is transcriptionally active in the absence of the SPI promoter. This result was confirmed using an in vivo transient expression assay system described in section 4.5.
4.2.1 Nucleotide Comparisons

Cattaneo et al., (1983b) originally proposed that sequences between HBV nucleotides 1188 to 1260 as the major promoter region for the HBsAg mRNA. This was based on the observation that these sequences, positioned immediately upstream of the mRNA initiation site, share extensive DNA nucleotide homology with sequences shown to regulate the transcription of the SV40 major late mRNA (Brady et al., 1982; Contreras et al., 1982; Cattaneo et al., 1983a). Examination of the transcriptional and genetic maps of all the sequenced Hepadnaviruses (see Chapter One, fig. 1.3 to 1.6 inclusive) indicates that all the hepatitis B like viruses encode a major surface antigen transcript in relatively the same position on their respective sequences. Therefore, it is possible that sequences important in regulation of transcription of the surface mRNAs could have been conserved throughout the evolution of the viruses.

The major HBV surface antigen promoter region and flanking sequences (from the unique Bst EII site to the end of the preS2 region, HBV130 positions 915-1436) were compared to the corresponding sequences of the other subtypes of HBV, GSHV and WHV sequences. Nucleotide comparisons were performed using the University of Wisconsin Genetics Group software package (v.4), gap and pretty programs. The output of these programs was manually edited to optimise alignment of the nucleotide sequences (fig. 4.2). It was not possible to conveniently align the surface promoter region of the DHBV sequence, due to its divergence from the sequences of the mammalian viruses.

The nucleotide comparisons indicate that there are blocks of conserved nucleotide sequence among the mammalian viruses. These blocks of nucleotide homology are shown on fig. 4.2 (a to d). Conserved DNA sequences may not necessarily represent transcriptional regulatory
Figure 4.2 Nucleotide comparison of the major surface promoter regions of the mammalian hepatitis B like viruses. The consensus sequence was generated by a plurality of 8 homology (ie. 8/10 matching nucleotides), presented below the sequences. Regions of conserved nucleotide homology are indicated (a to d). The initiation methionine codons for the preS1 and preS2 regions are indicated, the nucleotide positions of the mapped transcriptional initiation sites are indicated with a + above the respective sequence. The positions of relevant restriction endonuclease sites present in the HBV130 nucleotide sequence are shown.

(References: Galibert et al., 1979, 1982; Valenzuela et al., 1980; Fujiyama et al., 1983; Ono et al., 1983; Mandart et al., 1984; Seeger et al., 1984; Buscher et al., 1985; Enders et al., 1985; Kodama et al., 1985; Moroy et al., 1985; Okamaota et al., 1986; Pugh et al., 1986; Yaginuma et al., 1987b)
Plurality: 8.00

### PreSi

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HBV130</td>
<td>915</td>
<td>GGTcA.CCATATTTCTTGGGAACA.AGA.GCTAcA.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVadr</td>
<td>915</td>
<td>GGTcA.CCATATTTCTTGGGAACA.AGA.GCTAcACCAtggaggtttgt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVadr4</td>
<td>915</td>
<td>GGTcA.CCATATTTCTTGGGAACA.AGA.GCTAcAGCatggaggtttgt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVadrw</td>
<td>915</td>
<td>GGTcA.CCATATTTCTTGGGAACA.AGA.GCTAcAGCatggaggtttgt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVayr</td>
<td>915</td>
<td>GGTcA.CCATATTTCTTGGGAACA.AGA.GCTAcAGCatggaggtttgt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVayw</td>
<td>915</td>
<td>GGTcA.CCATATTTCTTGGGAACA.AGA.GCTAcAGCatggaggtttgt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSHV</td>
<td>926</td>
<td>GGTaAaCCATATTTCTTGGGAACACACAAGCCTagAGAaaataatgggca</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHV</td>
<td>929</td>
<td>GGTaAaCCATATTTCTTGGGAACACACAAGCCTagGCAacataatgggca</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHV2</td>
<td>925</td>
<td>GGTaAaCCATATTTCTTGGGAACACACAAGCCTagGCAacataatgggca</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td></td>
<td>GGT-A-CCATATTTCTTGGGAACA-AGA--GCTA-AGCA------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DHBV</td>
<td>1121</td>
<td>GGTcAgCcTtTaATGTTGGGAACA 1143</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### HBV130

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HBVadr</td>
<td>961</td>
<td>ctccaaaacc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVadr4</td>
<td>961</td>
<td>ctccaaaacc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVadrw</td>
<td></td>
<td>ctccaaaacc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVayr</td>
<td>961</td>
<td>ctccaaaacc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVayw</td>
<td></td>
<td>ctccaaaacc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSHV</td>
<td>976</td>
<td>acaacataaaagtaacattcgatcccaacacaaatagctgcatgtgggtgca</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHV</td>
<td>979</td>
<td>acaacataaaagtcacccctccatccagacaaaaatagcagcgtgtgggcct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHV2</td>
<td>975</td>
<td>acaacataaaagtcacccctccatccagacaaaaatagcagcgtgtgggcct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td></td>
<td>acaacataaaagtaacattcccttccttccagacaaaaatagcagcgtgtgggcct</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### HBV130

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HBVadr</td>
<td>971</td>
<td>tcgac.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVadr4</td>
<td>971</td>
<td>tcgac.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVadrw</td>
<td>956</td>
<td>tcgac.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVayr</td>
<td>971</td>
<td>tcgac.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVayw</td>
<td></td>
<td>tcgac.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSHV</td>
<td>1026</td>
<td>acagtgggaacacattatatcaccacccacaccgctgacaaatccag.ca</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHV</td>
<td>1029</td>
<td>gcagtgggaacacattatatcaccacccacaccctccctcag...aacagtcag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHV2</td>
<td>1025</td>
<td>gcagtgggaacacattatatcaccacccacaccctccctcag...aacagtcag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td></td>
<td>gcagtgggaacacattatatcaccacccacaccctccctcag...aacagtcag</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### PreSi

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HBV130</td>
<td>946</td>
<td>...gcatggggcaGAATtcTtCCaccacgCAacTCttgCttggaATtc.tttcC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVadr</td>
<td>976</td>
<td>aagggcatgggAcaGAATtcTtCtCttcCaatcCTctggaATtc.tttC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVadr4</td>
<td>976</td>
<td>aagggcatgggAcaGAATtcTtCtCttcCaatcCTctggaATtc.tttC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVadrw</td>
<td>961</td>
<td>aagggcatgggAcaGAATtcTtCtCttcCaatcCTctggaATtc.tttC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVayr</td>
<td>982</td>
<td>aagggcatgggAcaGAATtcTtCtCttcCaatcCTctggaATtc.tttC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBVayw</td>
<td>946</td>
<td>aagggcatgggAcaGAATtcTtCCaccacgCAacTCttgCttggaATtc.tttC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSHV</td>
<td>1075</td>
<td>ta.ttcacaccAGAATTTTCagaccaacgtctttTaaaATcCaaggaag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHV</td>
<td>1075</td>
<td>tgtttacaccAGAATTTTCaacaacaacatctttTgtagaaATCCaacaac</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHV2</td>
<td>1071</td>
<td>tgtttacaccAGAATTTTCaacaacaacatctttTgtagaaATCCaacaac</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td></td>
<td>acaacataaaagtaacattcgatcccaacaaatagctgcatgtgggtgca</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

A-GAAT-T-TC-----CA---CT-----AT-C-----C
HBV130 1179  CcaG...cAaatCcGCTcCtCtCAACATCccAgtcAgGAgCcA  
HBVadr 1212  CcaG...cAagccCCTcCtCtCtCtCAACATCGgcAgtcAgGAAGAcA  
HBVadr4 1212  CcaG...tAgaCCTCCTCtCtCtCtCAACATCGgcAgtcAgGAAGAcA  
HBVadv2 1244  tcaA...cAaaCCTCCTCtCtCtCtCtCAACATCGgcAgtcAgGAAGAcA  
HBVayr 1212  CcaG...cAagccCCTCCTCtCtCtCtCAACATCGgcAgtcAgGAAGAcA  
HBVayw 1179  CcaG...cAaatCCGCTcCtCtCtCtCAACATCGgcAgtcAgGAAGAcA  
GSHV 1308  CtgctGgcAgtcCCTCAGaCcCtctCCACATCGgAtcAaAGGAaA  
WHV 1305  CtctGgtttAgttCCGCTCAaaCtCcaAAACATCGAgAtcAaAGGAaA  
WHV2 1316  CtgcGGtttAgttCCGCTCAaaCtCcaAAACATCGAgAtcAaAGGAaA  
Consensus  C--G----A---CC---CCTC---C--C---ACCAATCG--A---AGGAAG--A

Saul

HBV130 1226  GCCTACCCGctGtGtTCCACTCCCTGgAAaCCAcACATCAACCTCTCggCAgtc  
HBVadr 1259  GCCTACCCCAaTctCTCCACCTCTCCTAAGAGAAGAATCCTCTCggCAgtc  
HBVadr4 1259  GCCTACCCCaTctCTCCACCTCTCCTAAGAGAAGAATCCTCTCggCAgtc  
HBVadv2 1244  GCCTACCCCaTctCTCCACCTCTCCTAAGAGAAGAATCCTCTCggCAgtc  
HBVayr 1259  GCCTACCCCaTctCTCCACCTCTCCTAAGAGAAGAATCCTCTCggCAgtc  
HBVayw 1226  GCCTACCCCaTctCTCCACCTCTCCTAAGAGAAGAATCCTCTCggCAgtc  
GSHV 1358  aCcaCaaCCtCtaaCTCCtCCTtAGAGACACaACCTCTCttcttAgACA  
WHV 1355  GCCTACCCCTCaaCTCCACCTCTAAGAGACATCCTCCTCAAACTActA  
WHV2 1366  GCCTACCCCTCaaCTCCACCTCTAAGAGACATCCTCCTCAAACTActA  
Consensus  GCCTAC--CC--T--CTCCACCTCT-AGAGAGA-TCA-CTCTA----A---

HBV130 1276  AGtggAAaACtccAaacccttCCaCC............AaactCTGCaAGAT  
HBVadr 1309  AGtggAAaACtccAaacccttCCaCC............AaactCTGCaAGAT  
HBVadr4 1309  AGtggAAaACtccAaacccttCCaCC............AaactCTGCaAGAT  
HBVadv2 1294  AGtggAAaACtccAaacccttCCaCC............AaactCTGCaAGAT  
HBVayr 1309  AGtggAAaACtccAaacccttCCaCC............AaactCTGCaAGAT  
HBVayw 1276  AGtggAAaACtccAaacccttCCaCC............AaactCTGCaAGAT  
GSHV 1408  tCaaaAAtcagACagtcagtcaggttatGcAgAgGagAAGGAGt  
WHV 1405  tCaaaAAtcagACatctctctCttCagggGttctggtGcAgAgGagAAGGAGt  
WHV2 1416  tCaaaAAtcagACatctctctCttCagggGttctggtGcAgAgGagAAGGAGt  
Consensus  -G----AA----AC--------C--CC------------A----CTGC-AGA-

HBV130 1314  CcaGAgAg tgAagGgcCcGcCtCtAttttcCTGGCtgcAATtcgAGGcAgGAC  
HBVadr 1347  CcaGAgAg tgAagGgcCcGcCtCtAttttcCTGGCtgcAATtcgAGGcAgGAC  
HBVadr4 1347  CcaGAgAg tgAagGgcCcGcCtCtAttttcCTGGCtgcAATtcgAGGcAgGAC  
HBVadv2 1332  CcaGAgAg tgAagGgcCcGcCtCtAttttcCTGGCtgcAATtcgAGGcAgGAC  
HBVayr 1347  CcaGAgAg tgAagGgcCcGcCtCtAttttcCTGGCtgcAATtcgAGGcAgGAC  
HBVayw 1314  CcaGAgAg tgAagGgcCcGcCtCtAttttcCTGGCtgcAATtcgAGGcAgGAC  
GSHV 1458  CtaAaccccttcGatCttCaaACTCCtCtGTAAGatccTTttactC  
WHV 1455  tGgaCaaAcAaGcCagCagCaccaATCggAtcTGAgagatctTTttacG  
WHV2 1466  tGgaCaaAcAaGcCagCagCaccaATCggAtcTGAgagatctTTttacG  
Consensus  C---A-A---A--G--C---A-----CTGC---TGG-----TT-----AC

---

159
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HBV130</td>
<td>1364</td>
<td>gGATGGGGACCTGcgCtGaac......</td>
<td>1436</td>
<td></td>
</tr>
<tr>
<td>HBVadr</td>
<td>1397</td>
<td>gGACtGGGGACCCCTGcaCcGaac....</td>
<td>1469</td>
<td></td>
</tr>
<tr>
<td>HBVadr4</td>
<td>1397</td>
<td>gGAcTGGGGACCCCTGcaCcGaac....</td>
<td>1469</td>
<td></td>
</tr>
<tr>
<td>HBVadw</td>
<td>1382</td>
<td>gGAcTGGGGACCCCTGcaCcGaac....</td>
<td>1454</td>
<td></td>
</tr>
<tr>
<td>HBVadw2</td>
<td>1403</td>
<td>gGAcTGGGGACCCCTGctgacGatc...</td>
<td>1454</td>
<td></td>
</tr>
<tr>
<td>HBVayr</td>
<td>1397</td>
<td>gGAcTGGGGACCCCTGctgacGaac...</td>
<td>1475</td>
<td></td>
</tr>
<tr>
<td>HBVayw</td>
<td>1364</td>
<td>gGAcTGGGGACCCCTGctgacGaac...</td>
<td>1475</td>
<td></td>
</tr>
<tr>
<td>GSHV</td>
<td>1508</td>
<td>AcTAAgCCCTGttctgaCTACTGtCTCcaCCAcATtGTCtctctccaTtGa...</td>
<td>1508</td>
<td></td>
</tr>
<tr>
<td>WHV</td>
<td>1505</td>
<td>AcTAAgCCCTGttctgaCTACTGtCTCcaCCAcATtGTCtctctccaTtGa...</td>
<td>1505</td>
<td></td>
</tr>
<tr>
<td>WHV2</td>
<td>1516</td>
<td>AcTAAgCCCTGctggttcCTACTGtCTCcaCCAcATtGTCtctctccaTtGa...</td>
<td>1516</td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td></td>
<td>A-TAA-CCCTG------CTACTG-CTC--CCATAT-GTC------CTCGA</td>
<td>1508</td>
<td></td>
</tr>
</tbody>
</table>

160
sequences. The nucleotide sequences studied here covers not only the preS coding sequences but also the overlapping polymerase gene. Blocks of conserved nucleotide sequences may also represent conserved coding sequences for functionally important overlapping amino acid translations. The nucleotide sequence may be conserved due to the coding pressure for the overlapping reading frames limiting sequence divergence.

Region (a) is very highly conserved among all the sequenced hepatitis B like viruses including DHBV. This block of conserved nucleotides is found within the polymerase coding sequence. It is possible that this nucleotide sequence could code for a functionally important amino acid sequence of the polymerase protein and therefore has been conserved. This region is situated at position -340 to -313 relative to the start of the major surface mRNA (Cattaneo et al., 1983b) and +8 to +35 relative to the start of the preS1 mRNA (Rall et al., 1983; Ou and Rutter, 1985). Region (a) is outside the region defined by Siddiqui et al., (1986), to contain the promoter sequences for the major surface mRNA, which correspond to HBV130 positions 993 to 1310. Therefore, it is possible, that since this region is close to the initiation point of the preS1 mRNA, that it may be involved in the regulation of the preS1 mRNA transcription. No preS1 mRNA initiation sites have been mapped for GSHV and WHV. However, Buscher et al., (1985) have mapped an initiation site for a preS mRNA of DHBV, at positions 1107 and 1115 on the DHBV map. Therefore the DHBV conserved region (a) is situated at position +14 to +36 and +6 to +28 relative to the start of the DHBV preS mRNAs, in a similar position of HBV region (a), relative to the HBV preS1 mRNA initiation site.

A 10 base sequence found between HBV nucleotides 1036 to 1046 is highly conserved among the mammalian hepatitis B like viruses, region (b). The central nucleotides 1039-1043 (ATTGG) found in all sequences, except WHV2 (AcTGG), is a potential consensus sequence for an inverted CTF (CCAAT transcription factor) binding site. Such an inverted sequence has been found to be important for transcription of the Herpes simplex thymidine kinase gene at position -86 to -82 relative to the start of HSV tk gene transcription. This indicates that transcriptional activity
of the CTF binding site (CCAAT) is independent of orientation (Jones et al., 1985; Graves et al., 1986). The pentanucleotide 5'-CCAAT-3' occurs in the -80 region of many eukaryotic mRNA-coding genes (Benoist et al., 1980; Efstratiadis et al., 1980), therefore, the position of the distal ATTGG sequence (-216 to -212) is in an unusual position relative to the HBV major surface mRNA start site, possibly casting doubt on its involvement in HBV transcription. The sequences of the other mammalian hepatitis B "like" viruses also contain additional consensus sequences for an inverted CTF binding site at positions 1180-1184 for GSHV and 1214-1218 for WHV. This means that all the mammalian viruses contain a potential inverted CTF binding site at approximately position -215 to -220 relative to the mapped mRNA initiation sites (Enders et al., 1985; Moroy et al., 1985). The DHBV also contains a inverted CTF consensus sequence at position 1134-1138, within region (a), at -224 relative to the surface mRNA initiation site mapped by Buscher et al., (1985). Since all the viruses contain an inverted CTF consensus sequence at approximately the same position relative to the major surface mRNA initiation sites, it is suggestive that this sequence could be involved in regulation of the surface mRNA transcription. An additional potential CTF consensus sequence is conserved among the HBV and GSHV viruses at HBV position 1204-1208 (CCAAT), within region (c), -51 relative to the mapped mRNA initiation site. The corresponding WHV sequence contains a single mismatch (CaAAT). This sequence is placed in a more usual position of CTF binding sites observed by Benoist et al., (1980) and Efstratiadis et al., (1980).

The sequences proposed by Cattaneo et al., (1983b) as being the promoter for the major surface mRNA, is highly conserved among the viruses, region (c), possibly indicating conservation of important promoter elements. An interesting feature of the nucleotide sequence immediately upstream of the preS2 initiation codon is the alternating blocks of pyrimidine and purine rich sequences. The semi-conserved sequence motif TC(C/A)NCCTC is repeated three times within region (c), once within the region with homology to the SV40 origin and twice within the alternating pyrimidine, purine blocks. It is possible that these sequences could represent multiple binding sequences for a
transcriptional regulatory factor similar to the multiple Sp1 binding sites found in the SV40 early promoter region (reviewed by Kadonaga et al., 1986).

Region (d) covers a large number of small blocks of conserved nucleotide sequences. However this region is at the beginning of a large stretch of highly conserved nucleotide sequences covering the surface gene coding sequence and may represent important conserved coding sequences of the preS2 reading frame. The highly conserved region (d') shows a high degree of nucleotide homology to part of the SV40 72 base pair repeat enhancer. Mutations within the first repeated TC motif (5'-TCCCCAG-3') in the SV40 enhancer have been shown to reduce enhancer function (Zenke et al., 1986). The SV40 repeated TC motif has the overall sequence 5'-GTCCCCAGGCTCCCCAG-3' (complementary sequence 5'-CT0GGAGccI.GGGAC - 3'). The HBV sequence between nucleotides 1418-1428 (5'-TGGGGACCCTG-3'), shares the underlined nucleotide homology to the repeated SV40 TC motif. Since this HBV sequence shares homology to part of a known enhancer element, it is possible that this sequence could be involved in the regulation of surface mRNA synthesis. However, it could also be a coincidence that this sequence shares homology to the SV40 enhancer, conservation of this sequence may simply represent the conservation of functionally important amino acid sequences of the preS2 protein.

4.2.2 The Nuclear Factor-1 Binding Site

While the above nucleotide comparisons were being conducted Shaul et al., (1986a) published that nuclear factor-1 (NF-1), isolated from a WISH cell nuclear extract, binds to the HBVadw2 surface promoter region and protects nucleotides 1103-1122 of HBVadw2 (1064-1083 of HBV130) from in vitro DNase I digestion. Shaul et al., (1986a) propose that the NF-1 protein binds to the HBV sequence TGG(N)4GCAA (HBV130 positions 1068-1081) (fig. 4.2). Shaul et al., (1986a) have compared the relative in vitro binding affinities of the HBVadw and HBVayw sequences for the NF-1 protein. Their results indicate that the HBVayw binding site is five
fold less efficient at binding NF-1 than the corresponding HBVadv2 NF-1 binding site. Shaul et al., (1986a) attribute the lower binding affinity to an intrinsic property of the HBVayw sequence. A CC dinucleotide immediately 5' to the NF-1 binding consensus sequence in the HBVayw sequence could reduce the binding affinity of the sequence for NF-1. This phenomenon was also observed for deletion mutants of the Adenovirus 2 NF-1 binding site (Leegwater et al., 1985). Another possible explanation could be that the lower binding affinity of the sequence 5'-CC*TGG(N)GCCAA-3' (complementary sequence 3'-GGACC(N)CGGTT-5') could be the result of steric hindrance due to overlapping dcm methylation (dcm methylation consensus sequence CC(T/A)GG underlined with the internal C residue being methylated), as a result of amplification of the cloned viral DNA in E. coli. De Vries et al., (1987) have shown that partial in vitro methylation of the guanine residues in the Ad2 NF-1 consensus binding site reduces the NF-1 binding affinity of the sequence. Therefore it is possible that the lower NF-1 binding affinity of the HBVayw sequence could be due to overlapping dcm methylation. Unfortunately, Shaul et al., (1986a) did not discuss the strain of E. coli used to amplify the cloned HBV sequences, therefore, the effect of overlapping dcm methylation on the binding affinity of NF-1 to the HBVayw binding site is only speculative. However, Leegwater et al., (1985) used E. coli HB101 as a host for their constructions. E. coli HB101 is dcm+ (see below) and this could account for the low NF-1 binding affinity of the sequence CCTGG(N)GCCAA used in their studies. Shaul et al., (1986a) did not test the in vitro binding affinity of the HBV subtypes HBVadr and HBVayr sequences for the NF-1 protein; both these sequences contain a single mismatch to the NF-1 consensus sequence (TGG(N)GCCAA). Therefore, it is not possible to say whether the single mismatch to the NF-1 consensus sequence will be tolerated or impair the affinity of NF-1 binding to the surface promoter regions of these HBV subtypes.

The report of Jones et al., (1987) indicates that the NF-1 and CTF are possibly the same, or closely related nuclear proteins. Therefore, all the putative HBV NF-1 binding sites contain a single mismatch to the CTF consensus sequence in the 3' portion of the "consensus" NF-1 binding
site (adw and ayw: CCAAc; adr and ayr: CaAAT). The 5' portions of the HBV NF-1 consensus sequences also contain potential inverted CTF binding sites, with the adr (AaTGG) and adw/ayw (AcTGG) sequences containing a single mismatch to the inverted CTF consensus sequence. The subtype adr4 (ATTGG) contains a perfect inverted CTF consensus sequence while the ayw/HBV130 (ccTGG) sequence contains two mismatches. If NF-1 and CTF are the same nuclear protein, the CC dinucleotide, immediately 5' to the HBVayw NF-1 consensus sequence, may as Shaul et al., (1986a) propose result in a decreased binding affinity of NF-1/CTF to the sequence, due to two mismatches to the consensus motif, reducing/preventing the nuclear protein from binding to the DNA. If this is the case the observed in vitro binding affinity of NF-1 to the HBVayw sequence may be an intrinsic property of the nucleotide sequence and not an artifact due to dcm methylation. The WHV and GSHV promoter regions contain additional potential inverted CTF binding sites ATTGG, in approximately the same position as the HBV NF-1/CTF binding site relative to the conserved SPII promoter sequences (see above). Conservation of at least two potential distal binding sites for the CTF protein in all the mammalian hepadnaviruses suggests that these sequences could be involved in the regulation of transcription of the major surface mRNAs.

The results of Gronostajski, (1987) indicate that the CTF motif (CCAAT) is not sufficient for strong binding of purified NF-1 to sequences containing this motif, suggesting that NF-1 and CTF may be closely related proteins but not identical. It is possible that CTF may bind to NF-1 sites but the reciprocal may not always be the case. This hypothesis was supported by Miksicek et al., (1987) who have shown that the NF-1 protein failed to show significant binding to the CTF binding site of the Herpes simplex virus tk promoter which lacks the full NF-1 consensus sequence, indicating that NF-1 and CTF are probably not the same nuclear protein.

The results of Shaul et al., (1986a) indicate that the HBVadw2 NF-1 binding site is involved in the regulation of the SPII promoter. Their constructions contained the HBV sequences corresponding to HBV130 positions 973 to 1310 inserted upstream of the bacterial chloramphenicol
acetyltransferase gene (CAT gene). Deletion of the the 5' sequences to the corresponding HBV130 position 1065 reduced CAT expression by approximately 40% relative to maximal expression directed by the sequences 973-1310. Deletion to this position removed the potential inverted CTF binding sequence (1039-1043) observed in the nucleotide comparisons (see above), possibly indicating that the inverted CTF binding site is active in regulation of HBsAg mRNA transcription. Although it is possible that deletion to this position also removed important 5' flanking sequences necessary for optimal NF-1 binding. Deletion to position 1071 (within the NF-1 binding site) reduced expression by approximately 90%, indicating that HBsAg mRNA transcription is dependent upon the NF-1 binding site. Both deletion mutants indicate that the sequences between the corresponding HBV130 positions 973 to 1071 contain distal transcriptional regulatory sequences for the major HBV surface antigen promoter.

4.3 E. coli Methylation of HBV Sequences

The HBV130 NF-1 sequence has the same 5' flanking sequences as the HBV\textit{ayw} NF-1 binding site. One possible explanation for the low binding affinity of the HBV\textit{ayw} sequence for NF-1, could be the result of steric hindrance due to overlapping \textit{dcm} methylation. Therefore, it may be possible to modulate HBV surface mRNA transcription both \textit{in vitro} and \textit{in vivo} by using cloned HBV DNA amplified in both \textit{dcm}+ and \textit{dcm}− E. coli hosts. This hypothesis would only be correct if the relative level of transcription/expression is proportional to the level of NF-1 binding to the promoter sequences. Leegwater \textit{et al.}, (1985) used \textit{E. coli} HB101 as a host for their Ad2 NF-1 deletion mutants, therefore the two \textit{E. coli} strains HB101 (\textit{dcm}+) and GM48 (\textit{dcm}−) were chosen as hosts to amplify cloned HBV DNA sequences for use in \textit{in vitro} run off transcription and \textit{in vivo} transient expression assays.

The \textit{E. coli} strains HB101 and GM48 were tested for expression of the \textit{dcm} methylase by bacteriophage λ plating efficiency assays using an \textit{E. coli} strain expressing \textit{Eco RII}. The restriction endonuclease \textit{Eco RII}
(5'-CC(A/T)GG-3') is very sensitive to methylation and will not cleave DNA at its recognition sequence when the internal C residue is methylated and therefore, DNA isolated from a E. coli (dcm+) host is partially resistant to digestion with Eco RII (Bigger et al., 1973). Since E. coli HB101 contains the mutation hsdR20, the DNA isolated from this strain will not be methylated at either Eco B or Eco K sites, therefore, it was necessary to construct a E. coli (rK−) host containing the plasmid R245. Plasmid DNA was isolated from E.coli NM145 (R245, Tc+, mK+, r−, Eco RIP'), by the alkaline-SDS method. The plasmid DNA was used to transform E. coli 5K; one tetracycline resistant transformant was designated 5K/R245 and was used for the bacteriophage λ plating assays. Efficiency of plating experiments using bacteriophage λr gr grown on HB101 (λr, HB101) and GM48 (λr, GM48) were conducted on E. coli 5K and 5K/R245. The results of these plating assays are presented in table 4.2. The results indicate that E. coli HB101 codes for a functional dcm methylase; λr, HB101 was partially protected from restriction by Eco RII when plated on 5K/R245. Whereas λr, GM48 was restricted when plated on strain 5K/R245 confirming that GM48 (dam3, dcm6) does not code for the dcm methylase. These results were confirmed in vitro using pHBV130 DNA isolated from HB101 and GM48 backgrounds. The restriction endonuclease Bst NI (5'-CC(A/T)GG-3') is an isoschizomer of Eco RII but cleaves dcm methylated DNA (fig. 4.3). Plasmid DNA isolated from HB101 was partially resistant to cleavage with Eco RII but sensitive to restriction with Bst NI, while DNA isolated from a GM48 background was sensitive to both restriction endonucleases indicating that the dcm sites were not methylated. DNA isolated from a GM48 background was also sensitive to Bcl I digestion (data not shown) indicating that the strain GM48 is also deficient for the E. coli dam methylase. The strain E. coli GM48 codes for an active Eco K restriction system; it is unlikely that methylation of DNA by the Eco K methylase would affect HBV transcription since the HBV sequence contained in pHBV130 is devoid of Eco K (and Eco B) recognition sequences.

The clone λLTK154-H7 grew very poorly on GM48 with very low titres and no significant amounts of λ DNA was obtained when propagating this clone on GM48. Therefore, the 5·6 kb Hind III insert contained in
# Table 4.2 Efficiency of Plating Assays

<table>
<thead>
<tr>
<th>Titre on 5K</th>
<th>Titre on 5K/R245</th>
<th>EOP of $\lambda_{\text{vir}}$ on E. coli 5K/R245 relative to 5K=1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>pfu/ml</td>
<td>pfu/ml</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{\text{vir}}$. HB101</td>
<td>2.4x10^8</td>
<td>1.2x10^7</td>
</tr>
<tr>
<td>$\lambda_{\text{vir}}$. GM48</td>
<td>2.0x10^9</td>
<td>1.9x10^4</td>
</tr>
</tbody>
</table>

Bacteriophage $\lambda_{\text{vir}}$ stocks were prepared on E. coli HB101 and GM48. Plating efficiencies were determined by titre on E. coli 5K and 5K/R245 (Eco RI) to test for partial protection from restriction by dcm methylation.
Figure 4.3 Restriction endonuclease digestion of dcm methylated and non-methylated pHBV130 DNA isolated from *E. coli* HB101 and GM48 respectively. pHBV130.GM48 (1 µg) plus 3 units of Eco RII (track 1), pHBV130.HB101 (1 µg) plus 3 units of Eco RII (track 2), pHBV130.GM48 (1 µg) & pHBV130.HB101 (1 µg) plus 6 units of Eco RII (track 3), pHBV130.GM48 (1 µg) plus 1 unit Bst NI (track 4), pHBV130.HB101 (1 µg) plus 1 unit Bst NI (track 5), pHBV130.GM48 (1 µg) & pHBV130.HB101 (1 µg) plus 2 units of Bst NI (track 6). All restriction endonuclease digestions were performed as recommended by the manufacturer, for 60 minutes. Size markers used were the 1 kb ladder (BRL) (track M), sizes are shown (in kb).
ALTK154-H7 was gel purified and ligated into the Hind III site of the vector λNM1149. The ligated DNA was used to transform E. coli NM154. A recombinant clone was isolated and found to contain the 5-6 kb insert; this clone, λ1149-H7, was used for all further assays of the HBV surface promoter.

4.4 In Vitro Transcription Assays

The original studies conducted to investigate the transcriptional regulatory sequences of the HBV surface antigen promoter(s) were those of Rall et al., (1983). Rall et al., (1983) conducted in vitro run off transcription assays using a HeLa whole cell extract (Manley et al., 1980). They observed transcription of the preS/surface open reading frame initiating at position 907±10, 25 nucleotides downstream of a "TATA" like sequence. Rall et al., (1983) failed to observe the major transcription initiation site(s) observed by 5' end mapping of in vivo surface mRNAs (Cattaneo et al., 1983b; Standring et al., 1984; Ou and Rutter 1985; Yaginuma et al., 1987b). The possible reason for the lack of observation of the major transcripts by Rall et al., (1983) is most likely due to the templates used for the run off transcription assays. These templates were discontinuous downstream of the Eco RI (1280) and the Bam HI (1310) sites, therefore, if the templates directed transcription from the major SPII promoter, the largest oligoribo- nucleotides generated would have been ca. 25 and 55 nucleotides respectively and possibly would not have been resolved on the gel electrophoresis system used.

The results obtained studying HBsAg production by the cell line RJK88/H7-1 were suggestive of transcription and expression of the HBsAg directed by the SPII promoter. However, the results obtained were not substantial enough to unequivocally show that the putative SPII promoter was responsible for expression of the surface mRNAs. Therefore, in vitro transcription experiments using HeLa whole cell extracts (Manley et al., 1980) were conducted in an attempt to show that the DNA fragment cloned in the recombinant λ1149-H7 (and ALTK154-H7) has the potential
to transcribe and express HBsAg. Plasmid (pHBV130) and bacteriophage λ (λ1149-H7) DNA was prepared from E. coli HB101 (dcm+) and GM48 (dcm−) backgrounds, to test the effect of dcm methylation of the sequence CCTGG (1066-1070) on the \textit{in vitro} binding of NF-1 to the surface promoter region, in an attempt to modulate \textit{in vitro} transcription.

The \textit{Hind} III to \textit{Pst} I (1800) fragment from the left hand end of the clone λ1149-H7 was gel purified. This DNA fragment contains the HBV sequences from 994 to 1800 and contains the sequences found by Siddiqui \textit{et al.}, (1986) (positions 993-1310) to encode for SPII promoter activity in transient expression studies conducted using CV-1 cells. A \textit{Bst} EII (915) to \textit{Pst} I (1800) fragment of pHBV130 was also gel purified as an additional control. Half the DNA (≈ 1.5 μg) present in the samples was digested with \textit{Bsm} I (1098) to delete the NF-1 binding site (1068-1081) from the SPII sequences (1188-1260).

Both sets of methylated and unmethylated DNA fragments (H7: 994-1800 and 1098-1800; pHBV130: 915-1800 and 1098-1800) were used for \textit{in vitro} run off transcription assays, as described in materials and methods. As an additional control, \textit{Pst} I digested SV40 DNA (BRL, supplied with the HeLa cell extracts) was used to show that the \textit{in vitro} transcription system was functioning as expected. The expected run off transcripts directed by the HBV DNA fragments should have approximate lengths of 545, 524 and 513 nucleotides, based on the map positions of the 5' ends of \textit{in vivo} transcripts (Yaginuma \textit{et al.}, 1987b). The products of the \textit{in vitro} transcription reactions were separated on a 3.5% (w/v) polyacrylamide/1.0% (w/v) agarose formaldehyde gel system. The gels were dried and exposed to Kodak XAR-5 X-ray film at -70°C. After one week exposure no signals corresponding to the expected HBV run off transcripts could be observed on the autoradiograph (data not shown).

The expected four SV40 directed run off transcripts (separated on a 1.3% (w/v) agarose/formaldehyde gel) of ca. 2.0, 1.9, 1.75 and 1.6 kb (measured against 5' end labelled DNA molecular weight standards) gave strong signals on an overnight exposure using the less sensitive Cronex 4 X-ray film (data not shown). The sizes of the observed SV40 run off transcripts correspond to the size of the SV40 (\textit{Pst} I) run off
transcripts observed by Handa et al., (1981). This result indicates that the in vitro transcription system was functioning as expected. The HBV in vitro transcription experiments were repeated with the same result. Since no in vitro run off transcripts were observed, it is not possible to determine if overlapping dc m methylation of the NF-1 binding site affects the in vitro transcription efficiency.

Possible explanations for the lack of transcriptional activity of the linear isolated SPII promoter sequences could be i) trace amounts of contaminates present in the gel purified DNA solutions may have inhibited the in vitro RNA polymerase II transcription system, ii) the HeLa whole cell extract may be lacking certain transcription factors necessary for transcriptional activity of the unusual non-TATA SPII promoter, or, that such factors if present may be in the incorrect ratios to promote transcription, or iii) the SPII promoter sequences contained on the linear fragments between nucleotides 994 and 1800 are not transcriptionally active in the in vitro transcription system.

The only other reported in vitro transcription analysis of the SPII promoter is that of Cattaneo et al., (1983b). They observed HBV directed transcripts initiating at HBV position 1255. The in vitro transcription reactions of Cattaneo et al., (1983b) differed in two respects to those reported here and by Rall et al., (1983). The cell free extracts used by Cattaneo et al., (1983b) were not prepared in accordance to the procedure of Manley et al., (1980) but using a modified procedure of Weigaertner and Keller (1981), in which in vitro transcriptional activity depends upon the addition of a cytoplasmic fraction to a nuclear extract. The second difference in the method of Cattaneo et al., (1983b) is that they did not generate run off transcripts from linear templates, but transcribed cloned HBV DNA present in an uncut plasmid and mapped the start of transcription by SI protection mapping. The plasmid used by Cattaneo et al., (1983b) was pSH2·6, which contains 111% of the HBV genome from nucleotides (Hae II) 899–3182/1–1280 (Eco RI) inserted in a SV40 based vector, containing SV40 sequences (5157–5243/1–832) with the surface gene in sense with the SV40 late promoter. Therefore, the plasmid pSH2·6 contains not only the HBV enhancer element but also the
SV40 enhancer, both of which could effect transcription from the SPII promoter.

The SV40 enhancer element is known to stimulate *in vitro* transcription of heterologous promoter/enhancer systems (Sassone-Corsi *et al.*, 1984, 1985). However, the SV40 enhancer is unlikely to be totally responsible for expression of the *in vitro* transcripts directed by pSH2-6, since Cattaneo *et al.*, (1983b) report that cleavage of pSH2-6 with *Bam* HI (1004, 2682), which separates the SPII promoter from the SV40 enhancer, did not disrupt transcription from the SPII promoter. This *Bam* HI fragment contains the SPII promoter, the surface antigen gene and the HBV enhancer element. The results of Chang *et al.*, (1987b) indicate that the HBV enhancer element has the potential to stimulate surface gene transcription *in vivo*. The results of Shaul *et al.*, (1985) also indicate that, although, the HBV enhancer shows a semi-tissue specificity, it can function in non-hepatocyte derived cells. Since the SV40 enhancer is active *in vitro*, the HBV enhancer may also be active in HeLa cell extracts. HeLa cell nuclear extracts are known to contain an E "like" binding factor, which binds to the HBV enhancer region at position 2460-2486 (E binding site), that has been postulated to be involved in HBV enhancer function (Shaul and Ben-Levy, 1987). Therefore, it is likely that the HeLa cell extracts used for *in vitro* transcription assays also contain the E "like" binding factor. The difference in the results obtained in this study and those of Cattaneo *et al.*, (1983b), could be due to the fact that the templates used here were devoid of the HBV enhancer element and contained only the SPII promoter. All the DNA templates used by Cattaneo *et al.*, (1983b), for *in vitro* transcription experiments, contain the HBV enhancer element in *cis* with the SPII promoter. Therefore, *in vitro* transcription of the SPII promoter may be highly dependent upon the *cis* acting enhancer sequences. If this hypothesis is correct, the *Bam* HI fragment (1004-2682), from either λ1149-H7 or pHBV130, should direct a run off transcript of ca. 1.4 kb, while the same DNA fragments digested with *Pst* I will not generate run off transcripts due to separation of the SPII promoter sequences from the HBV enhancer.
4.5 Transient Assay Expression of HBsAg Using Mouse L Cells

4.5.1 Preliminary Transient Assay Experiments

Preliminary transient assay experiments were conducted according to the combined procedures of Sompayrac and Danna, (1981) and Luthman and Magnusson, (1983). Mouse L cells (3x10⁵ cells per dish) were plated onto 5x3.5 cm diameter dishes and allowed to grow overnight at 37°C in a CO₂ incubator, to be slightly sub-confluent at time of transfection. The gel purified Hind III DNA fragment from λ1149-H7, and pHBV130 DNA (0.2 µg DNA per 10⁶ plating cells) was mixed with a total volume of 0.7 ml of 200 µg/ml DEAE-dextran (molecular weight, 5x10⁵), 0.05 M Tris-HCl pH 7.3 in DMEM (serum free). The cells were incubated with the DNA mixture for 6 hours at 37°C in a CO₂ incubator. Chloroquine was added to a final concentration of 100 µM and the incubation continued for 2 hours. The medium was replaced with DMEM (7% (v/v) FCS) and the cell allowed to grow at 37°C/CO₂. The medium and cells were harvested 24 hours post transfection (ca. 70-80% confluence) and total nucleic acids isolated from the cells. The medium was assayed for the presence of HBsAg using an AUSRIA-II radioimmunoassay (Abbott laboratories). No HBsAg could be detected using either neat medium or concentrated samples (160,000 g for 12 hours, pellet resuspended in 1/30 initial volume of PBS) (data not shown). Separation of the total nucleic acids on a 1.3% (w/v) agarose/formaldehyde gel, transfer to hybond-N filters and probing with ³²P-labelled genomic length HBV DNA, showed the presence of a DNAase sensitive nucleic acid homologous to HBV within the cells (data not shown). This indicated that the transformation procedure was successful, in that the cells had taken up the transforming DNA, however, no detectable HBsAg was produced.

Gough and Murray, (1982) have shown that the cell line Rat2/130.4/TK4 only secreted significant levels of HBsAg into the medium when the cells were above 80% confluence and showed maximal expression when the cells reached 100% confluence. Therefore, the transfection procedure was repeated and the cells allowed to grow to confluence before the medium
was harvested. The results of the radioimmunoassay for HBsAg are presented in table 4.3. The results indicate that HBsAg is produced and secreted from the cells at a low level and that the DNA sequences contained on the *Hind* III fragment cloned in *λ*1149-H7 are capable of directing transcription of the HBV surface antigen gene, thus confirming the results obtained studying HBsAg expression by the cell line RJK88/H7-1. The transient assay was repeated as described in materials and methods, using 10 cm diameter dishes, with pHBV130 DNA digested with *Bst* EII. The results of the radioimmunoassay (table 4.3), indicate that by using the larger dishes for the transient assays the relative level of expression of HBsAg was increased by more than 30-fold.

4.5.2 Transient assays for the Expression of HBsAg: Analysis of the Transcriptional Regulatory Signals

a) Construction of pHBV1004

As a control for the transient assays the plasmid pHBV1004 was constructed; pHBV1004 contains HBV sequences between the *Bam* HI (1004) restriction site and the *Bgl* II (84) restriction site ligated into the *Bam* HI site of the cloning vector pUC8. The subtype of HBV cloned into pHBV130 contains two *Bam* HI sites at position 1004 and 2682. Therefore, to construct the plasmid pHBV1004; 10 µg of the plasmid pHBV130 was digested to completion with 20 units of *Bgl* II, in 20 µl of a buffer consisting of 50 mM NaCl, 10 mM Tris-HCl pH 7.4, 6 mM MgCl₂, 6 mM β-mercaptoethanol, at 37°C for 60 minutes. The *Bgl* II digested DNA was further partially digested with 1 unit of *Bam* HI in the same buffer at room temperature for 30 minutes. The digestion reactions were terminated by phenol extraction and the DNA fragments separated on a 1% (w/v) low melting temperature agarose gel. The 2-26 kb partial *Bam* HI/complete *Bgl* II DNA restricted fragment, spanning HBV nucleotides (*Bam* HI) 1004-3182/1-84 (*Bgl* II), was isolated and ligated into the *Bam* HI site of pUC8. The ligated DNA was used to transform *E. coli* JM83; one of the resultant recombinant plasmids isolated was designated pHBV1004 (fig. 4.4). The plasmid pHBV1004 contains the complete
Table 4.3  Radioimmunoassays for HBsAg from preliminary transient assays

<table>
<thead>
<tr>
<th>DNA† fragment used</th>
<th>P/N§ neat medium</th>
<th>P/N† conc. (x30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H7 (Hind III)</td>
<td>1.4</td>
<td>9.1</td>
</tr>
<tr>
<td>pHBV130 uncut</td>
<td>1.0</td>
<td>1.8</td>
</tr>
<tr>
<td>pHBV130 (Bst EII)</td>
<td>1.0</td>
<td>1.5</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pHBV130 (Bst EII)</td>
<td>4.1</td>
<td>35.2</td>
</tr>
</tbody>
</table>

† DNA samples were originally amplified in E. coli GM48 (dcm<sup>-</sup>) prior to the transient assays.
§ P/N ratios calculated against neat medium from a mock transformed mouse L cell negative control (P/N=1.0).
† P/N ratios calculated against x30 concentrated negative control samples prepared in the same manner as the test samples.

Radioimmunoassays were conducted in duplicate (200 µl) using the Abbott AUSRRIA-II radioimmunoassay for HBsAg. Concentrated samples were prepared by ultra-centrifugation, 12 ml samples were centrifuged at 35,000 rpm for 12 hours in a MSE 6×14 ml rotor, the pellets were resuspended in 400 µl of phosphate buffered saline. A) preliminary transformation according to the modified procedures of Sompayrac and Danna, (1981) and Luthman and Magnusson, (1983) using 5×3.5 cm diameter dishes. B) repeated assay using the procedure described in materials and methods using 1×10 cm dish.
Figure 4.4 Restriction endonuclease cleavage map of pHBV1004. The heavy lines represent HBV sequences and thin lines represent pUC8 sequences. pHBV1004 contains the HBV sequences from 1004 (Bam HI) to 84 (Bgl II) cloned into the Bam HI site of pUC8. The location and direction of the surface and X open translational reading frames are shown as the thick inner arcs, the thin inner arcs represent the HBV specific mRNAs. The nucleotide positions of restriction endonuclease sites unique to the HBV sequences are shown relative to the viral genes and transcriptional regulatory signals (shaded blocks).
transcriptional unit for the major surface mRNA, including the "SV40-late-like" (SPII) promoter, the preS2/surface open reading frame, the HBV enhancer, the X open reading frame and the polyadenylation signals (Simonsen and Levinson, 1983; Siddiqui et al., 1986; Shaul et al., 1985).

b) Transformation of mouse L cells

Methylated pHBV1004 and unmethylated pHBV1004, λ1149-H7 DNA samples were prepared using E. coli HB101 and GM48 hosts respectively. The Hind III fragment from λ1149-H7 was gel purified. The purified H7 fragment and pHBV1004 DNA were digested with the following enzymes: H7, Eco RI (mouse sequences), Bsm I (1098); pHBV1004 Eco RI (pUC8), Bsm I and Sau I (Mst II; 1150, 1264). The purified H7 fragment was digested with Eco RI to remove most of the 5' mouse flanking sequences. An Eco RI restriction site is situated 34 bp upstream of the integration site (HBV position 994); digestion at this site will remove any potential mouse promoters thereby eliminating possible co-transcription of mouse and HBV sequences. The pHBV1004 DNA was digested with Eco RI to separate the HBV surface gene from any potential 5' cryptic promoters contained on the vector sequences. The restriction enzyme Bsm I digests HBV DNA at position 1098, thereby, separating the NF-1 binding site and the SPII promoter sequences. Sau I (Mst II) digests HBV DNA at positions 1150 and 1264, either side of the SPII promoter sequence (Cattaneo et al., 1983b). Therefore, digestion of HBV DNA with Sau I separates the surface gene from the putative SPII promoter, and was used to show there are no transcriptionally active promoter sequences downstream of the sequence postulated by Cattaneo et al., (1983b).

The uncut plasmid and linear DNA fragments were used for duplicate transient expression assays by transfection of mouse L cells using DEAE-dextran as described in materials and methods. The transformed cells from the duplicate transient assays were pooled and total nucleic acids isolated. The medium from the transfected cells was assayed for production of HBsAg using the Abbott AUSRIA II-125 radioimmunoassay.
The level of HBsAg expression was quantified by comparison to standards prepared from a HBV carrier sample (HBVAg 5.8x10^3 ng HBsAg/ml) diluted in medium from the mock transfected mouse L cell negative control.

The RIA results obtained for HBsAg expression levels from mouse L cell transfections conducted using Sau I digested DNA, as expected, indicate that all the transcriptional regulatory sequences are contained at, or upstream, of the Sau I site at position 1264, supporting the results of Cattaneo et al., (1983b). The RIA results, for HBsAg expression by cells transfected with Bsm I digested DNA are suggestive that separation of the NF-1 site from the SPII promoter sequences, does reduce the level of expression of HBsAg. However, examination of the HBV DNA (RNAase digested) present within the transformed cells by Southern blot hybridisation (fig 4.5) indicates that the level of intra-cellular Bsm I digested HBV DNA was less, relative to the level of the other restriction digested DNAs. Examination of the relative levels of hybridisation by densitometric scanning of the autoradiograph (data not shown) indicated that the relative levels of hybridisation of linear Eco RI and Bsm I digested DNA fragments is approximately proportional to the relative difference in HBsAg expression. Therefore, the only possible conclusion that can be drawn from the results of this experiment, is that when HBV sequences between position 1310–84 are present downstream of the SPII promoter element, separation of the distal upstream NF-1 binding site does not inhibit expression of the HBsAg and, therefore, it is possible that NF-1 binding to the distal transcriptional element, observed by Shaul et al., (1986a) is not obligatory for transcription of the surface gene when using transfected mouse L cells.

The restriction enzyme Bsm I is one of the few known type II restriction enzymes that recognises a nonpalindromic sequence (5'–NGCATTC–3' complementary sequence 3'–tNCGTAAG–5') generating a two base 3' protruding end. It is possible that a minor proportion of the Bsm I digested transforming DNA could have recircularised or formed concatemers in vivo due to annealing/ligation of the protruding 3' ends reconstituting a functional promoter sequence (Folger et al., 1982;
### Table 4.4 Radioimmunoassay of transient assay medium

<table>
<thead>
<tr>
<th>DNA fragment digestion</th>
<th>P/N (neat) relative to mouse L cell=1</th>
<th>concentration HBsAg ng/ml</th>
<th>Average HBsAg ng/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse L cell</td>
<td>1.0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>H7† (Hind III)</td>
<td>A 7.7</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>B 7.8</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>H7† (Eco RI)</td>
<td>A 8.5</td>
<td>1.3</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>B 7.0</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>H7† (Bsm I)</td>
<td>A 2.5</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>B 2.2</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>pHBV1004† (Eco RI)</td>
<td>A 18.2</td>
<td>5.4</td>
<td>5.85</td>
</tr>
<tr>
<td></td>
<td>B 20.1</td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td>pHBV1004† (Bsm I)</td>
<td>A 14.0</td>
<td>3.4</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>B 13.3</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>pHBV1004† (Sau I)</td>
<td>A 0.9 (−ve)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>B 1.0 (−ve)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>pHBV1004$ (Eco RI)</td>
<td>A 20.7</td>
<td>6.6</td>
<td>6.35</td>
</tr>
<tr>
<td></td>
<td>B 19.7</td>
<td>6.1</td>
<td></td>
</tr>
<tr>
<td>pHBV1004$ (Bsm I)</td>
<td>A 16.4</td>
<td>4.4</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>B 15.1</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>pHBV1004$ (Sau I)</td>
<td>A 1.2 (−ve)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>B 1.1 (−ve)</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

† unmethylated DNA prepared from a E. coli GM48 background
$ methylated DNA prepared from a E. coli HB101 background

Transient assay transformations were performed in duplicate (A & B), the medium from each transformation was assayed for the presence of HBsAg in duplicate using the Abbott AUSRIA II-125 radioimmunoassay for HBsAg. The P/N ratios are expressed as an average of the obtained cpm values from each assay relative to the mock transfected mouse L cell negative control. The concentration of HBsAg was determined from a standard curve constructed from HBV-sag carrier serum (5.8×10^3 ng HBsAg/ml) diluted in medium from the negative control and assayed under the same conditions.
Figure 4.5 Examination of extra-chromosomal HBV DNA isolated from transfected mouse L cells. Total nucleic acids were isolated from transfected mouse L cells, digested with RNAase (heat treated), separated on a 0.7% (w/v) agarose gel, transferred to nitrocellulose and probed with a $^{32}$P-labelled Bst EII (915) to Bgl II (84) HBV restriction fragment of pHBV130 covering the surface transcriptional unit. Mock transfected mouse L cells (track 1), mouse L cells transfected with the gel purified Hind III H7 DNA fragment (track 2), H7 DNA digested with Eco RI (track 3), H7 DNA digested with Bsm I (track 4), pHBV1004 (non-methylated) (track 5), pHBV1004 (non-methylated) digested with Eco RI (track 6), pHBV1004 (non-methylated) digested with Bsm I (track 7), pHBV1004 (non-methylated) digested with Sau I (track 8), pHBV1004 (methylated) (track 9), pHBV1004 (methylated) digested with Eco RI (track 10), pHBV1004 (methylated) digested with Bsm I (track 11), pHBV1004 (methylated) digested with Sau I (track 12). Size markers used were the 1 kb ladder (BRL), sizes are shown (in kb).
Wilson et al., 1982; Wake et al., 1984). No recircularised molecules could be observed on the southern blot (fig. 4.5), although such molecules may be present in amounts below detection levels.

Recircularisation of the Sau I digested DNA would not reconstitute a promoter sequence since this enzyme digests at positions 1150 and 1264, any recircularised molecules would be deleted for the SPII promoter sequences. Since the level of HBsAg expression is proportional to the observed level of $^{32}$P-labelled HBV DNA hybridisation to linear DNA isolated from the cells, this is suggestive that the linear templates were used for transcription of the surface mRNAs. Although, the possibility that transcription from recircularised Bsm I digested molecules accounts for the observed HBsAg expression can not be excluded. If recircularised Bsm I digested DNA fragments were used as transcriptional templates this would mean that the majority of the uncut and linear, Eco RI digested, pHBV1004 DNA molecules were transcriptionally inactive.

If in vivo recircularisation is an artifact of the expression system this could be tested by repeating the transformation assay using pHBV1004 DNA digested at the unique Xho I site (1409). Xho I cuts HBV DNA between the SPII promoter and the coding sequence for the surface open reading frame. If mouse L cells express significant levels of HBsAg when transformed with Xho I digested pHBV1004, this would indicate that the molecules have recircularised.

It is evident that the presence of the restriction endonuclease Bsm I in the transformation solution, inhibited the efficiency of transformation. A similar phenomenon has also been observed by Gusew et al., (1987); using the calcium-phosphate transformation procedure to transform FR3T3 rat cells, they observed that certain restriction endonucleases, used to linearise DNA, inhibit transformation/integration by 10-50 fold and that removal of such enzymes by phenol extraction, prior to transformation, restored transformation efficiency. Gusew et al., (1987) propose that restriction enzyme proteins remain attached to the ends of the linear DNA following digestion and inhibit the integration of the DNA molecules into the chromosomes of the cells. Their results indicate that
the presence of the proteins did not effect DNA uptake into the cells, but decreased the efficiency of integration.

The DEAE-dextran transfection procedure used here does not rely upon integration of the linear DNA molecules therefore, the observed lower level of BsmI digested DNA within the cells can only be due to either i) the presence of the DNA bound protein inhibiting DNA uptake into the cells or ii) the presence of the attached protein promotes preferential DNA degradation. The difference in the possible mechanisms of decreased transformation efficiencies observed by Gusew et al., (1987) and this study, could be due to the different transformation procedures used and that the observed effects could be due to two separate phenomena resulting from the presence of the DNA bound protein.

One-third of the total nucleic acid preparation (equivalent to 2/3 of a culture grown on a 10 cm diameter dish) was separated by electrophoresis on a 1:3% (w/v) agarose/formaldehyde gel. Analysis by northern blotting, using a 32P-labelled genomic length HBV probe, showed the presence of a smear of nucleic acids hybridising to HBV DNA. The smear began at the expected molecular size of the linear transforming DNA molecules (data not shown). Digestion of the total nucleic acid with DNAase (RNAase free), prior to electrophoresis, removed the smear confirming that it was due to the presence of HBV DNA, however no surface mRNA could be detected (data not shown). It is possible that the RNA present within the samples was degraded during the isolation procedure, however, mouse rRNA bands were clearly visible upon ethidium bromide staining of the gels. It is possible that the amount of HBV surface mRNA was below detection levels.
4.6 Discussion: HBV Sequences Involved in the Regulation of Transcription of the Surface Gene

The above experiments were originally designed to generate results in support of those of Shaul et al., (1986a), in that one or more distal transcription elements (between HBV nucleotides 994-1098) are required for efficient transcription of the major HBV surface mRNAs. However, the results obtained in this study indicate that the distal NF-1 binding site (1068-1081) and the potential inverted CTF binding site (1039-1043) are not required for efficient expression of the HBsAg. All DNA sequences necessary for transcription of the surface mRNA are located downstream of the Bsm I site at position 1098. Therefore, overlapping dcm methylation of the NF-1 binding site will have no effect on HBsAg production, as indicated by the HBsAg RIA results obtained by analysis of the culture medium after transformation of mouse L cells with methylated and unmethylated pHBV1004 DNA. Deletion of the sequences 5' to the Sau I site at position 1264, resulted in the complete abolition of HBsAg expression most probably resulting from the deletion of the "SV40-late-like" promoter sequences postulated by Cattaneo et al., (1983b). Therefore, the major surface antigen promoter is contained between nucleotides 1098 and 1264, covering the region of nucleotide conservation found in the promoter regions of the mammalian hepadnaviruses.

Cattaneo et al., (1983b) originally proposed that sequences between HBV nucleotides 1188-1260 correspond to the major promoter for the HBsAg mRNAs, due to observed sequence homologies between these HBV sequences and regions of the SV40 genome known to regulate late transcription (Brady et al., 1982; Contreras et al., 1982; Cattaneo et al., 1983a). The postulation that the HBV sequences correspond to the surface promoter is based on the assumption that these sequences represent the same, or closely related transcriptional regulatory sequences as the SV40 late promoter. Therefore, to take the analogy further, it would be reasonable to assume that the putative major surface promoter interacts with the cellular...
transcriptional machinery in the same, or closely related manner, as the SV40 late promoter.

May and co-workers have conducted an extensive analysis of the sequences immediately upstream of the major SV40 late mRNA initiation site, positions -53 to +7 relative to the major late initiation site. This sequence contains the promoter sequences defined by Brady et al. (1982). Ernoult-Lange et al. (1984) constructed a series of SV40 based plasmids deleted for the SV40 early promoter region, the origin of replication, the G-C rich 21 base pair repeats (Sp1 binding sites) and the 72 base pair enhancer repeats. Their constructions contained the -53 to +7 late region inserted in the sense orientation upstream of the T antigen gene (i.e. in the opposite orientation to the native configuration). Transfection of HeLa cells with the construction resulted in no expression of the T antigen. Only when one or both of the 72 bp repeats were inserted in the native conformation relative to the -53 to +7 region, was significant T antigen expression observed. These results indicate that the -53 to +7 late region, in the absence of the enhancer sequences, does not display any intrinsic promoter activity.

Omilli et al. (1987) have shown that deletion of the -53 to +7 region reduces SV40 late transcription 5-fold, with the 5' initiation sites becoming more heterogeneous with respect to the native pattern of transcription sites. Inversion of the -53 to +7 region restores the native level of expression, however, only one 5' initiation site was observed. Examination of the sequence indicated that inversion of the region resulted in the generation of a "TATA" box like sequence in sense with the enhancer elements, with the 5' mRNA initiation beginning 21-26 bases downstream of the "TATA" like sequence. Omilli et al., (1987) propose that the -53 to +7 late region may not be an absolute requirement for late promoter activity driven by the enhancer elements. They suggest that the -53 to +7 region acts by correctly fixing the initiation sites and therefore, limits the heterogeneity of the late initiation sites.

The results of Shaul et al. (1986a) also indicate that the HBV "SV40-late-like" surface promoter requires cis acting elements for efficient transcription. Shaul et al., (1986a) inserted a HBVadw2 surface promoter
fragment (HBV130 positions 978-1310) upstream of the CAT gene. The fragment contains the putative NF-1 binding site, the major surface promoter and the cap sites of the HBsAg mRNA. Transfection of HeLa cells with their constructions, resulted in CAT expression. Complete isolation of the SPII promoter sequences, by deletion of the distal NF-1 binding site, reduced expression by 90% (from 2.9% to 0.3-0.4% conversion of chloramphenicol; or from 3.2 to 0.33-0.44 relative to pSV2CAT=100 (90% conversion)). The plasmid pSV2CAT contains the SV40 72 bp repeat enhancer, the SV40 early region promoter and the start site for transcription (Gorman et al., 1982). The relative level of CAT expression directed by the isolated surface promoter indicates that the isolated SPII promoter is transcriptionally weak when compared to expression directed by the SV40 early promoter contained on the plasmid pSV2CAT, when assayed under the same conditions.

It should be noted that in a previous report, Shaul et al., (1985) cloned a Bam HI fragment of HBVsdw2 (corresponding to nucleotides 2682-3182/1-1310 of HBV130), into the Bgl II site of pSVOCAT to generate pHB2.0CAT with the SPII promoter in sense with the CAT gene. Transfection of Alexander cells with this construction resulted in no significant expression of the CAT gene (<0.1 relative to pSV2CAT=100). The construction of Shaul et al., (1986a) contains the same HBVsdw SPII sequences (corresponding to HBV130 sequences 978-1310) in sense with the CAT gene. In this case transfection of HeLa cells with this construction resulted in a low, but significant level of CAT expression (see above). If all the SPII promoter regulatory sequences are contained upstream of position 1310 it would be expected that both constructions used by Shaul et al., (1985) and Shaul et al., (1986a) should show SPII directed CAT expression. A possible explanation for the lack of SPIII directed expression of the CAT gene in Alexander cells could be that the SPII promoter is not functional in these cells. This is clearly not the case; the Alexander cells are known to express the HBsAg particles and examination of the surface promoter directed HBV mRNA isolated from the cells indicates that ca. 98% of the surface mRNAs are directed by the SPII promoter (Ou and Rutter, 1985). Therefore the SPII promoter contained on pHB2.0CAT, should have directed CAT expression in Alexander.
cells if all the transcriptional regulatory sequences are contained 5' to position 1310. This is, unless the SPII promoter contained on the plasmid is being repressed by unknown HBV or cellular directed factors or that the SPII promoter contained on pHB2.0CAT was non-functional due to deletion or mutation.

Chang et al., (1987b) using a similar construction, containing HBVadw sequences (corresponding to HBV130 positions 523-1280) inserted 5' to the CAT gene (pSpCAT), conducted transfections of the hepatoma cell line HuH-7. Again, SPII promoter directed expression of the CAT gene was observed. In their assays the levels of conversion of chloramphenicol to its acetylated products, using the pSpCAT transformed HuH-7 cell lysates, was 4.06% conversion of chloramphenicol. Chang et al., (1987b) also inserted the HBV enhancer region 2245-2682 upstream of the surface promoter fragment to generate pEnSpCAT ([Hinc II) 2245-2682 (Bam HI)/Egl II) 523-1280 (Eco RI)]. Transfection of the hepatoma cell line with pEnSpCAT resulted in approximately 20-fold increase of the surface promoter directed CAT expression (90.54% conversion of chloramphenicol for pEnSpCAT transfected HuH-7 cell lysates). The level of CAT expression using constructions containing the HBV enhancer and the surface promoters, was of approximately the same order of magnitude as the level of CAT expression using the plasmid pSVwCAT (97.93%) assayed under identical conditions.

The isolated surface promoters have been studied by Shaul et al., (1986a), Siddiqui et al., (1986) and Chang et al., (1987b) by inserting HBVadw (HBVadw2) fragments containing the surface promoters into pSVOCAT or pSP6-CAT (Shaul et al., 1986a) based plasmids. In all cases, transfection of mammalian cells with the constructions containing the isolated surface promoter(s), resulted in CAT expression. The combination of these result indicates that transcriptionally weak SPII promoter regulatory sequences are found between HBV nucleotides 993 and 1310. It should be noted that all these investigations were carried out using subtype HBVadw sequences. It is possible that the weak transcriptional activity observed using these constructions may be unique to the HBVadw subtype. Shaul et al., (1986a) have shown that the
SPII promoter is highly dependent upon the HBVadw2 NF-1 binding site. They have also shown that the subtype HBVayw has a lower in vitro binding affinity for NF-1 relative to HBVadw2 and as Shaul et al., (1986a) propose, this could be an intrinsic property of the DNA sequence. The remaining subtypes of HBV all have a single mismatch to the NF-1 consensus sequence and it is possible that NF-1 will not bind to these degenerate sequences. Therefore, all HBV subtypes, other than HBVadw, may not contain independent transcriptionally active promoter elements on the isolated sequences between the corresponding HBV130 sequences 993 and 1310. The above expression studies need to be repeated using DNA sequences from subtypes HBVadr or HBVayr before the observed promoter activities of HBVadw can be generalised to include all subtypes of HBV.

Siddiqui et al., (1986) have constructed HBV-pSVOCAT recombinants containing the isolated "TATA" box promoter (SPI) or the SPII promoter. Transfection of human, non-human, hepatoma derived and non-hepatoma tissue culture cells with both sets of constructions, resulted in CAT expression. An interesting point made by Siddiqui et al., (1986) is that both the SPI and SPII promoters contained in their pSVOCAT derived plasmids, appear to have similar promoter strengths. This is in contrast to the actual relative promoter strengths found in Alexander cells, HBV infected liver tissue or hepatoma derived cell cultures supporting viral replication, in which the SPI directed mRNA accounts for only a minor proportion (approximately 2%) of the total surface antigen mRNA (Ou and Rutter, 1985; Will et al., 1987; Yaginuma et al., 1987b).

Summation of the above published results suggests that while the HBV sequences 993-1310 do contain elements of the SPII promoter, which under certain conditions are capable of independent transcriptional activity, they also suggest that not all the transcriptional regulatory elements are contained upstream of position 1310. The results of Chang et al., (1987b) indicate that the HBV enhancer may also regulate the expression of the SPII promoter. Siddiqui et al., (1986) suggest (based on unpublished results) that the HBV enhancer element, in its native
position relative to the surface gene, may also be involved in regulation of the SPII promoter in hepatocellular carcinoma derived cell lines.

It is possible that the difference in the results obtained by Shaul et al., (1986a) and the results obtained in this study, could simply be due to the different HBV sequences contained in the plasmid constructions and the interaction of these sequences, or HBV protein products coded by these sequences, with the SPII promoter. The constructions of Shaul et al., (1986a) did not contain the HBV enhancer or X gene sequences. The promoter activity observed by Shaul et al., (1986a) was highly dependent upon the upstream distal transcriptional elements contained on their constructions. Deletion of the potential inverted CCAAT box and the NF-1 binding site, both of which resulted in a decrease in promoter activity, resulted in the isolation of the "SV40-late-like" promoter sequences and like the actual SV40 late region (-53 to +7) these sequences contain no or very little intrinsic promoter activity in the absence of other cis acting elements. The constructions used in this study contained the complete transcriptional unit for the native major surface mRNAs (assuming no distal elements are contained upstream of position 1004) as well as the complete potential transcriptional unit for the X gene mRNAs. It is possible that the HBV enhancer element contained on the constructions used here has the ability to interact in cis with the SPII promoter sequences, having a much stronger influence on the SPII promoter than the upstream distal transcriptional element observed by Shaul et al., (1986a). The physical separation of the 5' distal elements from the SPII sequences by restriction endonuclease digestion (assuming that the DNA fragments did not recircularise in vivo), had very little, if any, affect on the overall expression of the HBsAg. However, the methods used to detect HBsAg expression do not differentiate between the relative levels of expression of the preS2 and major surface proteins. It is possible, that during natural HBV infection, the distal promoter region may be involved in regulation of transcription of the relative ratios of the preS2 and surface mRNAs, thereby regulating viral morphogenesis.
It is possible that the HBV "SV40-late-like" SPII promoter, like the actual SV40 late promoter, is highly dependent upon the influence of the viral enhancer sequences for full transcriptional activity. The HBV enhancer has been shown to be semi-tissue specific in its ability to stimulate core promoter activity, in that its stimulative effect was 50 fold higher in Alexander cells as compared to its activity in CV-1 cells (Shaul et al., 1985). It could be argued that due to this tissue specificity, the possible enhancer augmentation of the SPII promoter would be minimal in mouse L cells. However, Shaul et al., (1985) have also shown that in an heterologous system, the HBV enhancer element is capable of stimulating the SV40 early promoter activity, in the absence of the 72 bp repeat, 105-fold in COS cells, relative to the vector plasmid, pA1OCAT3, containing only the SV40 early promoter and the CAT gene. These results indicate that while the enhancer element may have maximal activity in hepatocyte derived cells, the enhancer has the potential to have a very significant stimulative effect on promoters in non-hepatocyte derived cells. Therefore, it is probable that the HBV enhancer may also have a significant stimulative effect on the SPII promoter in mouse L cells.

The results of Wollershiem and Hofschneider (1987) indicate that the X gene product may act as a trans-activating protein. The constructions used in the transient transformation assays also contain the potential complete X gene transcriptional unit. Therefore, it is possible that the X protein could have been expressed in the transformed cells either by translation from its own mRNA or by internal initiation of translation of the surface mRNA. If the X protein was translated, it could also have a trans-activating stimulative effect on the HBV SPII promoter.

4.8 Future Prospects

The sequences necessary for the regulation of transcription of the major surface mRNAs require further investigation. Although the results in this study indicate that the upstream distal element(s) proposed by Shaul et al., (1986a) are not required for transcription, when the sequences downstream of position 1310 are present, the results obtained
could be artifacts of the system due to possible recircularisation of the linear DNA fragments used, reconstituting a functional promoter sequence. Unfortunately, time did not permit further investigation of the sequences necessary for expression of the HBsAg. However, using the transient expression system described in this report with a series of deletion derivatives of the plasmid pHVB1004, it should be possible to determine whether the HBV enhancer element or the X protein influence the promoter activity of the SPII promoter. A series of deletion derivatives of the plasmid pHVB1004 could be constructed using unique restriction endonuclease sites present in the HBV sequences, but not in the pUC8 vector. An outline of the proposed construction procedure is presented in fig. 4.6.

The proposed constructions would be deleted for the following sequences from the surface transcriptional unit: pHVB1004ΔNF-1 (Sma I (pUC8) to HBV position 1098, Bsm I), pHVB1004ΔSPII (Sau I, 1150 to 1264), pHVB1004ΔEn (Dra III, 2318 to 2654, Nco I), pHVB1004ΔX (Nco I, 2654 to 3164, Sty I), and pHVB1004ΔEn-X (Dra III, 2318 to 3164, Sty I). The NF-1 binding site could also be removed from the constructions containing deletions of the enhancer and X gene by removing the Sma I to Bsm I fragment in the same way as creating the pHVB1004ΔNF-1 plasmid, to generate double deletion mutants. Transient assays could be performed in mouse L cells and possibly also in the HBV negative human hepatoma cell line HuH-7 (Nakabayashi et al., 1982), which has been shown to be a good host for both HBsAg expression and viral replication (Chang et al., 1987b, Yaginuma et al., 1987b). The results of the transient assays should indicate the interaction of the NF-1 binding site, the enhancer element and possibly the X protein with the SPII promoter and therefore increase our understanding of viral gene expression.
Figure 4.6 Proposed procedure for construction of deletion derivatives of pHBV1004: pHBV1004ΔNF-1; pHBV1004ΔSPII; pHBV1004ΔEn; pHBV1004ΔX; pHBV1004ΔEn-X. Thick lines represent HBV sequences, thin lines represent pUC8 sequences. Restriction endonuclease abbreviations: B, Bam HI; D, Dra III; E, Eco RI; N, Nco I; S, Sau I; Sm, Sma I; St, Sty I. For nucleotide positions of the restriction endonuclease sites relative to the translational open reading frames and transcriptional regulatory signals refer to figure 4.4.
pHBV1004

1. Smal/BsmI
2. T4 DNA polymerase
3. T4 ligase

D

1. Sau I
2. T4 ligase

St

1. DraI/NcoI
2. T4 DNA polymerase
3. T4 ligase
4. Gel purify

NcoI

1. Klenow DNA polymerase
2. T4 DNA polymerase
3. EcoRI
4. Gel purify

pHBV1004ANF-1

pHBV1004ASPII

pHBV1004AN

pHBV1004ANf

pHBV1004AX

pHBV1004AXf
Appendix I

Nucleotide Sequence of HBV130

The nucleotide sequence presented in this appendix is that contained in the clone pHBV130 (Pasek et al., 1979; Gough and Murray, 1982; Pugh et al., 1986). The nucleotide sequence shown is the sequence of the complete plus strand (3182 nucleotides) numbered using the convention of Pasek et al., (1979). Shown below the sequence is the amino acid translation of the major open translational reading frames found in the three forward reading frames, the positions of the methionine initiation codons for the viral genes are indicated. The mapped transcriptional initiation sites are indicated by + above the sequence (Treinin and Laub, 1987; Yaginuma et al., 1987b). Also shown are the positions of the polyadenylation signal and the direct repeats DR1 and DR2.
<table>
<thead>
<tr>
<th>Nucleotide Sequence of HBV130</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Core Poly(A)</td>
</tr>
<tr>
<td>ATGGCATGACCCTTTATAAAAGAATTGAGCTACTGTGAGTTACTCTC MetAspIleAspProTyrLysGluPheGlyAlaThrValGlueLeuLeuSe</td>
</tr>
<tr>
<td>51</td>
</tr>
<tr>
<td>GTTTGGCTCTGGGCTTCTTCTTCTCTCGGAGATCTCTTCTTAGATACCG rPheLeuProSerAspPhePheProSerValArgAspLeuLeuAspThrArg</td>
</tr>
<tr>
<td>101</td>
</tr>
<tr>
<td>CCGCAACTGCTGGTATCCGAGCTTCTGAGCTTCTCTGAGCTGAGCTCGGAGACTTAGCTAATCAlaAlaLeuTyrArgAspAlaLeuGluSerProGluHisCysSerPro</td>
</tr>
<tr>
<td>151</td>
</tr>
<tr>
<td>TCTAGCTCTGGTGCAGTTTTCTACTTATTTAAGAAGAATCTTCTGCTGCTGCTGGAGACTTAAATAGAGAGACTTTACTAC</td>
</tr>
<tr>
<td>201</td>
</tr>
<tr>
<td>TTCAGTTGTTCACTTTTCTTTCTTTGAGAAAGAAGCCTTTCTTAGALeuTrpPheHisIleSerCysLeuThrPheGlyArgThrValLeuGl</td>
</tr>
<tr>
<td>300</td>
</tr>
<tr>
<td>TAGTAGTCAGTGGTTATGCACACTACTTGGTGCTAAATTGCAGACTTAATCAeuValValSerTyrValAsnThrAsnValGlyLeuLysPheArgGlnLeu</td>
</tr>
<tr>
<td>400</td>
</tr>
<tr>
<td>CACCACATGCCCATATCTATCACACTTCCGGAGACTACTGCTTTCTAGAroProAsnAlaProIleLeuSerThrLeuProGluThrValValArg</td>
</tr>
<tr>
<td>500</td>
</tr>
<tr>
<td>CGACGATGGTGCCTAGAAGAAGAACTCCTCTGCTGCTGGAGACTCCACCAGGAG</td>
</tr>
<tr>
<td>550</td>
</tr>
<tr>
<td>600</td>
</tr>
<tr>
<td>AGATACCTTGGACTCTAAGGGGAAATTTTACTGGCTTTATTCTC SerIleProTrpThrHisLysValGlyAsnPheThrGlyLeuTyrSerSe</td>
</tr>
<tr>
<td>651</td>
</tr>
<tr>
<td>TACTGTACTGCTTTAACCCTCATGGGAATCTGCTTTTCTTAATA</td>
</tr>
<tr>
<td>700</td>
</tr>
<tr>
<td>rThrValProValPheAsnProHisTrpLysThrProSerPheProAsnI</td>
</tr>
<tr>
<td>651</td>
</tr>
<tr>
<td>TACATTACACAAAGATATTATCAAAAAATGTGAAACAGTTTGAGGGCCA</td>
</tr>
<tr>
<td>195</td>
</tr>
</tbody>
</table>
CTCACAGTCAATGAGAAAAGAAGGTTAAAATTGATCATGCCTGCTAGGTT
LeuThrValAsnGluLysArgArgLeuLysLeuIleMetProAlaArgPh
TTATCCTAATTTTACCAAAATATTGGCCTTGGATAAGGGTATTTAACCTT
eTyrProAsnPheThrLysTyrLeuProLeuAspLysGlyIleLysProT
ATTATCAGAACATCTAGTTAATCATTACTTCCAAACTGACACTATTTA
yrTyrProGluHisLeuValAsnHisTyrPheGlnThrArgHistyrLeu
CACACCTCTATGGAGGCGGTTGTTTTATATAAGAGATCTCAGACATAG
HisThrLeuTrpLysAlaGlyValLeuTyrLysArgValSerThrHisSe
+  ProSl
CGCCCTATTTTGGGTACCATATTTCTGGGAACAAGACGTACAGCATG
rAlaSerPheCysGlySerProTyrSerTrpGluGlnGluLeuGlnHisG
Met
GGGCAGAATCTTTTCCACGACATCTCTGAGATTCTTCTCCGACCACCA
lyAlaGluSerPheHistyrGlnGlnSerGlyIleLeuSerArgProProGlyGlnAsnLeuSerThrSerAsnProLeuGlyPhePheProAspHistyr
GlyGluSerLeuGlnSerGlyValGluValLeuGlnSerProAlaPheArgAlaAsnThrAsnAsnProAspTrpAspPheA
ATCCCCAAAGGACACCTGGCCAGACCCCAACAGGAGCTGGAGCAGCA
nSerGlnGlnGlyHisLeuAlaArgGlnGlnGlyArgSerTrpSerIysProAsnGlyAspThrTrpProAspAlaAsnGlyValGlyAlaAla
TTCCGGCTAGGGTCTACCCCCACGCGGAGGCCTTTTTGGGTGGAGCCC
leArgAlaArgValHisProThrAlaArgArgProPheGlyValGluPhePheGlyLeuGluPheThrProHisGlyGlyLeuGlyLeuTrpSerPr
TCAGGTCTAGGAGGATAACTGCAAACTTTGCGAGCAATCCGCCCTTGCC
SerGlySerGlyHisAsnAlaAsnLeuAlaSerLysSerAlaSerCysLe
oGlnAlaGlnGlyIleMetGlnThrLeuProAlaAsnProProAlaS
1201  CTACCAATCGCCAGTCAGGACGGCGACGTACCCCCGCTGTTCTCCACCTCTG
uTyrGlnSerProValArgThrAlaAlaTyrProAlaValSerThrSerG
erThrAsnArgGlnSerGlyArgGlnProThrProLeuSerProLeuProLeu
AGAACC ACTCATCCTCAGGCAATGGAACACCTCCACACAGAATGGAATGGT

luAsnHisSerSerSerGlyHisAlaValGluLeuHisAsnLeuProPro
ArgThrThrHisProGlnAlaMetGlnTrpAsnSerThrThrPheHisGl

AUCTCTGCAAGATCCAGACTGAGGAGGCTGTATTATTCCCTCTGCTGGTGCT

AsnSerAlaArgSerGlnSerGluArgProValPheProCysTrpTrpLe
ntThrLeuGlnAspArgValGlyLeuThrPheProAlaGlyGlyS

CCAGTTCAGGGACAGTAAACCCCTGCTCGACTCGCTCCCCCATATCG

uGlnPheArgAspSerLysProCysSerAspTyrCysLeuSerHisIleV
erSerSerGlyThrValAsnProValProThrThrAlaSerProIleSer

Surface

TCAATCTCTCGAGGATGGGACCTGCGCTGAACATGGAGAACACATCAC

alAsnLeuLeuGluAspArgPheCysAlaGluHisGluHisSerIlePheSerArgIleGlyAspProAlaLeuAsnMetGluAsnIleTh

ATCAGGATTCTAGGACCCCTGCTCGTGTTACAGGCGGGGTTTTTCTTGT

IleArgIleProArgThrProAlaArgValTyrGlyValPheProValVa
rSerGlyPheGlyLeuGlyGlyProLeuValGlnAlaGlyPhePheLeuL

TGACAAGATCCTCAATACGACAGACTGACGATCGGTGGGACTTCT

1AspLysAsnProHisAsnThrAlaGluSerArgLeuValAspPheSe
uThrArgIleLeuThrIleProGlnSerLeuAspSerThrTrpThrSer

CTCAATTTTCTAGGGAACCTACGCTGTCTGGGACATGGGACATTGCA

erGlnPheSerArgGlyAsnTyrArgValSerProLysPheAlaVal
LeuAspAspLeuAspPheThrGlyThrThrValCysLeuGlyGlnAsnSerGl

CTCAATTCAGGGAACCTACGCTGTCTGGGACATGGGACATTGCA

CCCAATCTCAAATCTACAACCAATCCCTGCTCCCTCAACTTGTTGCT

ProAsnLeuGlnSerLeuThrAsnLeuLeuSerSerAsnLeuSerTrpLe
rProIleSerAsnHisSerSerCysProProThrCysProGlyT

ATCGCTGGATGTGCTGCGGCGTTTTATCATCTTCTTCTTCTATCCTGCT

uSerLeuAspValSerAlaAlaPheTyrHisLeuProLeuHisProAlaA
yrArgTrpMetCysLeuArgArgPheIlePheLeuPheIleLeuLeu

CTATGCCTCATCTTCTTCTTCTTCTTCTGACTATCAAGTATGTTGCC

1aMetProHisLeuLeuValGlySerSerGlyLeuSerArgTyrValAla
LeuCysLeuIlePheLeuLeuValLeuAspTyrGlnGlyMetLeuPr
1751  .  .  .  .  1800
CGTTTGTCCTCTAATTCCAGGATCATCAACCACCAGCACGGGACCCTGCA
ArgLeuSerSerAsnSerArgIleIleAsnHisGlnHisGlyThrLeuGluValCysProLeuIleProGlySerSerThrThrThrThrGlyProCysA
1801  .  .  .  .  1850
GAACCTGACGACTCCTGCTCAAGGAATCTCTATGTATGATATCCCCCTCTGTGCC
nAsnLeuHisHisCysSerArgAsnLeuTyrValSerLeuLeuLeuLeuLrgThrCysThrThrProAlaGlnGlyIleSerMetTyrProSerCysCys
1851  .  .  .  .  1900
TGTAAAAACCTCCGGATGGGAAACTGCACCTGTATTTCCATCCATC
euTyrLysThrPheGlyTrpLysLeuHisLeuTyrSerHisProIleIle
CysThrLysProSerAspGlyAsnCysThrCysIleProIleProSerSe
1901  .  .  .  .  1950
CTGGCTTTTCGAAGAATTCTATGGGAGTGCTGTTCTGTAGCT
LeuGlyPheArgLysIleProMetGlyValGlyLeuSerProPheLeu
TrpAlaPheGlyLysPheLeuTrpGluTrpAlaSerAlaArgPheSerT
1951  .  .  .  .  2000
GGCTCAGTTTTACTGCTCAATTTGCTGTAGCT
uAlaGlnPheThrSerAlaIleCysSerValValArgAlaProHrpLeuSerLeuLeuValProPheValGlnGlyLeuSerPro
2001  .  .  .  .  2050
ATGGTTTGCTTTCAGTTATGAGATGTTGATTTGGGCTTCC
isCysLeuAlaPheSerTyrMetAspAspValValGlyAlaLysSerIleValTrpLeuSerValIleTrpMetMetTrpTyrTrpGlyProSerLe
2051  .  .  .  .  2100
GTACACATCTTTCGATCCCTCTTCTACCCTGTTACCAATTCTTTTTC
ValGlnHisLeuGluSerLeuPheThrAlaValThrAsnPheLeuLeuSeuTyrSerIleLeuSerProPheLeuProLeuProIlePhePheCysL
2101  .  .  .  .  2150
TTTGGGCATACATTTAAACCTAACAAAAACAAAAAGATGGGCTTATTCTC
rLeuGlyIleHisLeuAsnProAsnThrLysThrArgTyrGlyTyrSerLeuTrpAlaTyrIleEnd
2151  .  .  .  .  2200
TAATTTTCATGGCATTGCTATTGGATTTGGGATCATTCGACCAAGAT
euAsnPheMetGlyTyrValIleGlyCysTrpGlySerLeuProGlnAsp
2201  .  .  .  .  2250
CACATCACAACAAAATCCAAAGATTTTGGAAACTCTTGGTACAG
HisIleIleHisLysIleLysGluCysPheArgLysLeuProValHisArg
2251  .  .  .  .  2300
GGCATTGGAATGTTGCTACAGGATTTGGGCTTCTCTGTTTGT
gProIleAspTrpLysValCysGlnArgIleValGlyLeuLeuGlyPheA
CTGCCCCTTCACACAATGTGTTATCTGTATTTAATGCCTTTGTATGCA
laAlaProPheThrGlnCysGlyTyrProAlaLeuMetProLeuTyrAla
TGTATTCAATCTAAGCGGCTTTCACTTTTTTCGCCAACACTACAAGGCTT
CysIleGlnSerLysGlnAlaPheThrPheSerProThrTyrLysAlaPh
tCTGTTGAACAATACCTGACACCTTTTACCCCGTGCCAGGCAACGGCCAG
eLeuCysLysTyrLeuAsnLeuTyrProValAlaArgGlnArgProG
GTCTGTGCAAGTGGTCTGATGCAACCCCCACTGCTGGGCTTGTC
lyLeuCysGlnValPheAlaAspAlaThrProTyrGlyTrpGlyLeuVal
ATGGGCCATCACGCGTCTGGGAACCTTCTGTGCTCTTGCGCATCCA
MetGlyHisGlnArgMetArgGlyThrPheLeuAlaProLeuProIleHi
+ TACTGCCGAACCTCCTAGGCCCTTTTGCTGAGGCGAAGCAA
sThrAlaGluLeuLeuAlaAlaCysPheAlaArgSerArgSerGlyAlaA
+ ACATTTCGGAACCTGCAACTCTCTTGCTCTCCTCCGCAATATACATCG
snIleLeuGlyThrAspAsnSerValLeuSerArgLysTyrThrThrSer
X TTTCCATGCCTCCTAGGCTTGGCCACACTCACCTCCTGGGGACGT
MetAlaAlaArgLeuCysGlnLeuAspProAlaArgAspValPheProTrpLeuLeuGlyCysAlaAlaAsnTrpIleLeuArgGlyThrThrSer
+ CTTTTTTAAGTTGGCTGCTGGGCTGATGCGACCCCTTCTCGGG
LeuCysLeuArgProValAlaGluSerCysGlyArgProPheSerGly
rPheValTyrValProSerAlaLeuAsnProAlaAspAspProSerArgG
+ GCCGCTTGGGGCCTGTCTGGGCTTGGGCTGCCTGGGGCGGCC
yProLeuGlyAlaLeuSerSerSerLeuSerLeuProAlaValProAlaAlaH
lyArgLeuGlyProCysArgProLeuCysLeuSerProLeuPheProArgPro
2801 ACGGGGCGACCTCCTTCTTTACGGGGCTTCCCCCTGCTGGGGCTTACCT
isGlyAlaHisLeuSerLeuArgGlyLeuProValCysAlaPheSerSer
ThrGlyArgThrSerTyrAlaSerProSerValProSerHisLe
2851 DR2 GCCGGACCTTGCTGACCTCCTGCCAGTGGCAACGACC
A AlaGlyProCysAlaLeuArgPheThrSerAlaArgArgMetGluThrTh
uProAspArgValHisPheAlaSerProLeuHisValAlaTrpArgProP
2901 CGTGAACGCCCAATCTTGGCAGAAGACTTCTGT
tValAsnAlaHisGlnIleLeuProLysValLeuHisLysArgThrLeuG
GACTCTCTGCAATGTCAACGACCGACCTTGAGGCATACTTCAAAGACTGT
lyLeuSerAlaMetSerThrThrAspLeuGluAlaTyrPheLysAspCys

TTGTTAAAGACTGGGAGGAGTTGAGGAGAGATTAGATTAAGGCTTT
LeuPheLysAspTrpGluLeuGluGluGluIleArgLeuLysValPh

TGTACTAGGAGGCTGGGATAAAATTGCTGCGACCAGCACCAGGAGGCA
eValLeuGlyGlyCysArgHisLysLeuValCysAlaProAlaProCysA

MetC1

ACTTTTTCACCTCTCCTAATCATTCTCCTGCTACCTACTGTCAAG
snPhePheThrSerAlaEnd

nLeuPheHisLeuCysLeuIleIleSerCysSerCysProThrValGlnA

CCTCAAGCTGTGGCCTTGGGTGGCTTTGGGGC

laSerLysLeuCysLeuGlyTrpLeuTrpGly/MetAspIle....(Core)
Bibliography


Gross-Ballard, M., Oudet, P. and Chambon, P. (1973). Isolation of high-
molecular-weight DNA from mammalian cells. *European Journal of

Gusew, N., Nepveu, A. and Chartrand, P. (1987). Linear DNA must have free
ends to transform rat cells efficiently. *Molecular and General
Genetics* 206:122-125.

Transcription of Simian virus 40 DNA in a HeLa whole cell extract.
*The Journal of Biological Chemistry* 256:478-482.

Hart, R. P., McDevitt, M. A. and Nevins, J. R. (1985) Definition of
essential sequences and functional equivalence of elements down-
stream of the adenovirus E2A and the early simian virus 40
polyadenylation sites. *Molecular and Cellular Biology* 5:2975-2983.

cellular onc gene by promoter insertion on ALV-induced lymphoid

Heermann, K. H., Goldrnann, V., Schwartz, W., Seyffarth, T., Baumgarten, H.
virus contain the pre-S sequence. *Journal of Virology* 52:396-402.

Herring, A. J., Inglis, N. F., Ojeh, C. K., Snodgrass, D. R. and Menzles, J. D.
(1982). Rapid diagnosis of rotavirus infection by direct detection
of viral nucleic acid in silver-stained polyacrylamide gels. *Journal
of Clinical Microbiology* 16:473-475.

Hino, O., Kitagawa, T., Koihe, K., Kobayashi, M., Hare, M., Mori, N.,
hepatitis B virus DNA in hepatocellular carcinoma in Japan.
*Hepatology* 4:90-95.

Hino, O., Shows, T. B. and Rogler, C. E. (1986). Hepatitis B virus
integration site in hepatocellular carcinoma at chromosome 17;18
translocation. *Proceedings of the National Academy of
Sciences. U.S.A.* 83:8338-8342

Virus-like particles in sera of patients with infectious and serum


Hohn, B. (1979). *In vitro* packaging of \( \lambda \) and cosmid DNA. *Methods in

bacteriophage particles *in vitro*. *Proceedings of the National Academy

209


