REVIEW ARTICLE

The Biology of Hepadnaviruses

By COLIN R. HOWARD

Department of Medical Microbiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, U.K.

INTRODUCTION

The family hepadnaviridae at present consists of seven hepatotropic viruses with double-stranded DNA genomes, all of which may cause persistent infections in their natural hosts (Table 1). Hepatitis B virus of man (HBV) was for many years the sole virus recognized in this group. Its identification followed the chance discovery of the so-called 'Australia antigen' (Blumberg et al., 1965) and the subsequent demonstration of the association of this antigen with viral hepatitis B (Blumberg et al., 1968; Prince, 1968; Okochi & Murakami, 1968). Consequently, this virus has received considerable attention, particularly as an estimated 200 million people are persistently infected with HBV. Many of these individuals are predisposed towards the development of hepatocellular carcinoma, the most common cause of neoplasia in many parts of the developing world (Zuckerman, 1982). However, the lack of conventional tissue culture systems has hampered the study of the replication and transformation potential of these viruses.

The finding of related hepadnaviruses in naturally infected woodchucks, ground squirrels and Pekin ducks has provided valuable animal models which, together with the availability of human hepatoma cell lines, has rapidly broadened our horizons as to hepadnavirus replication and genome expression. Despite these advances, prevention of hepatitis B in man still remains a major undertaking. Only the eradication of the virus from areas where maternal and perinatal transmission occur frequently will result in a significant reduction in the numbers of HBV carriers and the susceptibility of these individuals to the development of primary liver cancer later in life. This review describes recent information on the replication of the hepadnaviruses, with emphasis on how the newly discovered members of non-human origin are aiding our understanding of the complex pathogenesis of these agents, in particular the virus of human hepatitis B which represents the most significant chronic virus infection of man.

Epidemiology

Serological surveys have shown the existence of HBV carriers in all countries of the world. However, there are marked differences in the prevalence of hepatitis B virus surface antigen (HBsAg), the outer surface coat of the virus shed into the circulation from the hepatocytes of individuals infected with HBV (Fig. 1). The prevalence of HBsAg in adults ranges widely from 0.01 to 0.1% in northern Europe, North America and Australia to approximately 5% in countries bordering the Mediterranean, parts of eastern Europe, the Middle and Far East, and to 15% or more in tropical areas. Even within the same area there are wide variations between different ethnic groups, and even between villages a few miles apart (Whittle et al., 1983). Information regarding the distribution of the remaining hepadnaviruses is scarce, although almost certainly other members of the group await discovery. Those identified so far have been isolated from a wide variety of vertebrate species (Table 1).

Perhaps the most remarkable epidemiological feature of HBV infection in man is the incubation period, which may extend from 2 to 6 months before the development of clinical disease. Shorter incubation periods have been reported, however, and there may be some overlap within the range associated with the unrelated hepatitis A virus. Much has been written regarding the transmission, diagnosis and prevention of HBV and will not be dealt with here.
Fig. 1. Electron micrograph of the human hepatitis B virus (HBV) and related morphological forms present in the sera of persistently infected individuals. The 42 nm double-shelled virus contains an outer surface coat which bears common antigenic determinants with the 22 nm spherical and filamentous particles (HBsAg). The virus contains a distinct 27 nm core which in its morphologically intact state bears the core antigens (HBcAg). Disruption of the nucleocapsid results in release of the third antigenic component (HBeAg) which may also exist free in the sera of those individuals with high levels of HBV. Bar marker represents 50 nm.

Table 1. Comparative properties of the hepadnaviruses

<table>
<thead>
<tr>
<th>Host</th>
<th>Abbreviation</th>
<th>% Persistent infections</th>
<th>Hepatocellular carcinoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man</td>
<td>HBV</td>
<td>0.01–25</td>
<td>Yes</td>
</tr>
<tr>
<td>Woodchuck</td>
<td>WHBV</td>
<td>16–30</td>
<td>Yes</td>
</tr>
<tr>
<td>Ground squirrel</td>
<td>GSHV</td>
<td>0–50</td>
<td>?</td>
</tr>
<tr>
<td>Duck</td>
<td>DHV</td>
<td>1–10</td>
<td>?</td>
</tr>
<tr>
<td>Tree squirrel</td>
<td>TSHV*</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Stink snake</td>
<td>SSHV*</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Kangaroo</td>
<td>KHV*</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

* Unpublished information.
Review: The hepadnaviruses

For extensive reviews, the reader is referred to Deinhardt & Deinhardt (1983), Vyas et al. (1984) and Tiollais et al. (1985). With the introduction of sensitive tests for the markers of HBV, it has become increasingly apparent that modes of transmission must exist other than direct inoculation of blood and blood products. Viral antigen has been detected in saliva, semen, breast milk, amniotic fluid and in various tissue and body fluids contaminated by blood. Close personal contact may therefore be of primary importance in the spread of HBV, although there is no evidence to suggest aerosol transmission per se. In tropical and Asian regions, the epidemiological pattern is quite different from that in North America, Europe and Australasia. Most persons become exposed to the virus during childhood, either from carrier siblings or from direct contact with their carrier mother.

Pathogenesis and serological properties

Electron microscopy of infectious human serum shows the presence of a 42 nm particle (Fig. 1). Other hepadnaviruses are morphologically identical, with the exception of the duck hepatitis virus (DHV) which has a larger overall diameter and larger inner core component or nucleocapsid. The antigenic determinants associated with the inner nucleocapsid are termed collectively as HbcAg in the HBV system.

The hepatitis B virus core antigen (HbcAg) is first detected in the serum of an infected person on average 4 weeks prior to clinical or laboratory evidence of liver damage and may persist in most cases until the onset of symptoms and liver dysfunction (Fig. 2). Antibody to the core component is found during the acute phase of the disease and the presence of anti-Hbc is a marker of recent or ongoing virus replication. It has been suggested that recovery from HBV infection is accompanied by a relatively short-lived anti-HBs response, while a normal immune response of greater longevity is produced against HbcAg. The strong association of this antibody response with the persistence of HbsAg suggests that anti-Hbc at low titre in the population indicates that HBV infection may frequently be inapparent and probably transient. Carriers of HBV with a high titre of circulating virus often show a sustained IgM response to HbcAg.

The e antigen system represents the third antigenic marker of HBV and free HBeAg is often found in the circulation of patients during the early stages of infection and in patients with
chronic active hepatitis. The appearance of anti-HBe correlates with a good prognosis and a
decline in virus replication. The presence of HBeAg in the blood of persistently infected
individuals is a further indication of high levels of HBV and correlates with a higher titre of
HBsAg. All evidence obtained so far indicates that HBeAg is structurally related to the HBV
core and is encoded by the C gene; treatment of HBCAg-positive particles with detergents
abolishes HBCAg reactivity with the release of low mol. wt. HBeAg (Mackay et al., 1981).
Hepatocytes supporting high levels of virus replication therefore appear to shed into the
circulation large quantities of this additional structural polypeptide which frequently becomes
associated with serum proteins (Tedder & Bull, 1979). HBCAg may be regarded as representing
mainly conformational antigenic determinants resulting from the assembly of the inner core
component.

In the majority of naturally occurring cases of HBV infection, HBsAg is most likely to be
detected during the first week of the acute phase of illness and may persist from a few days to
several weeks. Continual expression of HBsAg beyond 6 months is generally taken as indicating
the development of a carrier state, and the antigen has been continually detected in the serum of
some persons over periods as long as 20 years. Anti-HBs generally develops some weeks or
months after recovery, albeit at a low titre detectable only by sensitive assay methods such as
radioimmunoassay or passive haemagglutination. The reason for this is not clear, although
impairment of B cell response to HBsAg determinants may occur during acute infection.

The presence of large quantities of antigen during antigenaemia make detection of anti-HBs
difficult during the later stages of acute HBV infection although anti-HBs can occasionally be
detected as circulating antigen–antibody complexes. A high anti-HBs titre may occur when
there is a history of repeated exposure to the antigen, often in the absence of clinical disease. In
parallel with the development of a humoral response to HBsAg, specific cell-mediated immunity
has been demonstrated (Edgington & Chisari, 1975; Thomas et al., 1985). The positive cell-
mediated response appears to be transient, beginning 2 to 3 months after the onset of disease and
is more or less simultaneous with the clearing of HBsAg from the circulation. Cytotoxic T cells
are suspected of playing a role in the immunopathology of acute hepatitis B but this has yet to be
demonstrated owing to the lack of a suitable target cell for assays in vitro.

A mild attack of acute hepatitis B may predispose an individual to the development of severe
HBsAg-associated chronic hepatitis. This may develop in approximately 10% of patients
admitted to hospital with acute icteric type B hepatitis. In one study, a third of these patients
suffered from chronic active hepatitis, displaying a spectrum of hepatic lesions and sporadic
episodes of jaundice (Redeker, 1975). The remaining two-thirds showed signs of persistence of
HBsAg but were otherwise in good health. In the latter group, resolution of chronic persistent
hepatitis may occur over 1 to 3 years, although serum HBsAg persists. Comparison of HBsAg
titres between the two groups has shown a significantly higher titre of circulating HBsAg in
cases of chronic persistent hepatitis.

In general terms, the pathology of HBsAg-associated chronic aggressive hepatitis closely
resembles the clinical syndrome of active chronic hepatitis in which 15 to 20% of patients
possess circulating HBsAg. However, of the remaining HBsAg-negative active chronic hepatitis
cases, over 60% are found to possess a significant delayed hypersensitivity response to HBsAg,
suggesting that a past exposure to HBV may have represented an important event in the
development of chronic liver disease. This is further indicated by clinical observations of
patients progressing from HBsAg-positive acute to HBsAg-negative active chronic hepatitis.
Immunofluorescence has frequently been used to study the distribution of HBV gene products in
vivo. HBCAg reactivity appears often by this technique to be restricted to the nucleus and
perinuclear region of the infected hepatocyte, whereas HBsAg is confined to the cytoplasm.
During the early stages of acute hepatitis, both reactivities are present in liver tissue. HBsAg is
typically distributed in a diffuse fashion in the cytoplasm of hepatocytes throughout the liver.

The presence of the HBeAg antigenic moiety distinct from the HBs/HBc antigen systems
appears to predispose the patient to the development of chronic liver disease. HBeAg is often
present in the sera of persistent carriers of HBsAg on haemodialysis, but only infrequently
among chronic carriers in the donor populations. A close relation has been shown between
HBeAg and the presence of HBcAg. Carriers of HBsAg negative for HBeAg with low levels of HBV show no histological or biochemical signs of liver disease. Hence, the presence or absence of HBeAg may be an indication of the extent of HBV activity and the subsequent course of the disease after infection.

It has been noted that the pathological changes seen in animals infected with other hepadnaviruses show important differences compared to HBV in man. For example, ground squirrels infected with ground squirrel hepatitis virus (GSHV) show little evidence of hepatitis (Marion & Robinson, 1983). Woodchucks chronically infected with woodchuck hepatitis B virus (WHBV) often show signs of moderate to severe hepatitis, but the development of cirrhosis as seen in man has not been found (Popper et al., 1981). These differences appear unrelated to the level of virus in the blood, as infected ground squirrels often show very high concentrations of virus (Marion et al., 1980) and indeed HBV infection in chimpanzees may also result in virus levels far in excess of human cases of acute HBV, despite the relatively minor histological abnormalities in infected non-human primates (Zuckerman et al., 1979). The inference from these observations is that liver cell injury is immune-mediated rather than a direct cytopathic effect of hepadnavirus infection. Given the range of viral antigen expression and genome activity in different cells of the same biopsy specimen, the latter cannot be entirely ruled out.

Virus has been recovered from cells removed from infected tissues but as yet no in vitro system exists for the study of hepadnavirus replication. Transfection experiments of yeast, mouse cells and other cell lines using cloned DNA allow the expression of at least some gene products, notably the S gene, in a morphological form which closely resembles the 22 nm HBsAg particle seen in the sera of individuals and demonstrates that broadly different eukaryotic cell types of non-primate origin may support at least some of the stages of virus replication. The critical stages of virus adsorption and genome penetration have yet to be mimicked in vitro, however, although early experiments using cultured hepatocytes would suggest these events are amenable to study in the laboratory (Zuckerman, 1975). It is worth considering the possible reasons for these difficulties. The hepatocyte is a highly differentiated parenchymal cell and the maintenance in vitro of the highly specialized functions of this cell type may be important for supporting the full replication cycle of these viruses. Alternatively, a second cell type may be required for presenting the virus to the hepatocyte. Mims (1982) has pointed to the role of Kupffer cells lining the liver blood vessels in determining the hepatotropism of viruses introduced into the circulation of the host. Certainly, the long incubation period of HBV in man contrasts markedly with the shorter incubation period of hepatitis A virus and indeed many other viruses which infect the liver. This observation together with the peak of viraemia prior to or at the onset of the clinical disease would tend to suggest that the hepatocyte is not the preferred primary cell type of these viruses although this hypothesis remains to be thoroughly investigated.

**Structure and organization of the hepadnavirus genome**

The product of the endogenously primed DNA polymerase reaction has been characterized as DNA with an approximate sedimentation coefficient of 15S. The observation that the reaction was not stimulated by the addition of a number of synthetic or natural DNA structures indicates that the reaction template was sequestered within the core of the 42 nm virus particle (Kaplan et al., 1973). Several studies have since shown that a high proportion of hepatitis B virions separated from serum positive for HBsAg contain the DNA genome. Robinson et al. (1974) visualized circular molecules with a mean length of 0.78 μm after extraction of core particles with SDS and 2-mercaptoethanol. The size of the circular forms was compatible with an estimated mol. wt. of 1.6 × 10^6 to 2.0 × 10^6, considerably smaller than the genome of any other known double-stranded DNA virus.

The hepadnavirus genome appeared at that time to be smaller than the total amount of nucleic acid required to encode for all the polypeptides bearing virus-specific antigenic determinants expressed during acute and persistent infection. It has since been determined from nucleotide sequencing studies, however, that the reading frames of several HBV gene products overlap and that the many polypeptides associated with HBsAg-bearing particles result from post-
Fig. 3. Genetic organization of the HBV genome. Both HBsAg (S gene) and HBcAg (C gene) are gene products of the complete positive DNA strand and may be translated together with precursor sequences pre-C and pre-S. A large reading frame (P gene) capable of transcription and translation into a 93000 mol. wt. protein may represent the virus-associated polymerase, but this remains unproven. The shorter strand overlaps the ends of the L strand to form an almost double-stranded circle of DNA. Two transcripts have been found. The larger transcript is read over the whole molecule and may represent the mRNA of the C and putative P genes. The shorter transcript is translated to produce the pre-S and S gene products, and possibly the X protein.

translational modifications or initiation of translation at various sites on the genome upstream to the major S gene. Delius et al. (1983) have carefully measured the extent and position of the single-stranded gap. This was variable with a preferred minimum length of about 700 nucleotides, i.e. approximately 22% of the total genome, in more than 99% of the molecules examined. By contrast, the genome of DHV is mostly fully double-stranded (Summers & Mason, 1982), an observation which may be of relevance in extrapolating the deduced sequence of events during replication of the duck hepadnavirus to HBV. Delius and colleagues suggest that these single-stranded regions are rich in potential sites for hairpin formation and may play a role in DNA packaging and virus assembly. This would infer a more direct role in virus replication for incomplete double-stranded structures than the suggestion of Ruiz-Opazo et al. (1982) that they represent defective genomes.

The apparently unique structure of HBV DNA and associated polymerase prompted Summers et al. (1975) to apply restriction enzyme cleavage techniques to synthetic DNA prepared by the reaction of Escherichia coli DNA polymerase I with denatured viral DNA as the template. The high specific activity product was resolved by gel electrophoresis and found to consist largely of a homogeneous 3600 nucleotide component, together with small quantities of a more rapidly migrating heterogeneous component containing a total of 3132 nucleotides. Comparison of these two species confirmed that the DNA polymerase activity associated with the intact 42 nm virus particle may function as a repair enzyme by closing single-stranded regions in the otherwise double-stranded structure.
The hepadnaviruses

Table 2. Nucleotide sequence studies of hepadnavirus genomes

<table>
<thead>
<tr>
<th>Virus</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human (100)*</td>
<td>Serum, subtype adw</td>
<td>Burrell et al. (1979)</td>
</tr>
<tr>
<td></td>
<td>Serum pool</td>
<td>Valenzuela et al. (1979)</td>
</tr>
<tr>
<td></td>
<td>Subtype adw</td>
<td>Sninsky et al. (1979)</td>
</tr>
<tr>
<td></td>
<td>Subtype ayw</td>
<td>Galibert et al. (1982)</td>
</tr>
<tr>
<td></td>
<td>Subtypes adr and adw</td>
<td>Ono et al. (1983)</td>
</tr>
<tr>
<td></td>
<td>Subtype adr</td>
<td>Fujiyama et al. (1983)</td>
</tr>
<tr>
<td>Woodchuck (60–70)</td>
<td>Serum pool</td>
<td>Galibert et al. (1982)</td>
</tr>
<tr>
<td>Pekin duck (&lt;50)</td>
<td>Source pool</td>
<td>Mandart et al. (1984)</td>
</tr>
<tr>
<td>Ground squirrel (50)</td>
<td>Source pool</td>
<td>Seeger et al. (1984)</td>
</tr>
</tbody>
</table>

* Figures in parentheses show the extent of homology (expressed as a percentage) of each virus compared with HBV of humans (100%).

Table 3. Major genes and products of the hepadnavirus genome

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product (mol. wt.)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>25000 and 30000 glycoproteins*</td>
<td>Major envelope protein, bearing HBsAg determinants</td>
</tr>
<tr>
<td>C</td>
<td>17000–22000†</td>
<td>Internal core protein, bearing conformationally dependent HBcAg determinants; also exists in serum as antigenically distinct e antigen</td>
</tr>
<tr>
<td>P</td>
<td>65000 (estimate)</td>
<td>Virus-associated DNA polymerase?</td>
</tr>
<tr>
<td>X</td>
<td>17000–28000†</td>
<td>5' end linkage to negative strand?</td>
</tr>
</tbody>
</table>

* Larger products present in minor amounts resulting from expression of additional pre-S determinants (see text).
† C and X gene products expressed as one polypeptide in duck hepatitis B virus.

At least six complete HBV nucleotide sequences are available as a result of cloning of the native DNA structures. The restriction enzyme maps obtained by these workers have shown some similarities and differences between DNA cloned for sources containing different HBsAg subtypes (Burrell et al., 1979; Sninsky et al., 1979; Charnay et al., 1979; Valenzuela et al., 1979; Fujiyama et al., 1983), although differences between preparations expressing HBsAg of the same subtype have also been reported (Marion et al., 1980). Table 2 lists the data available on HBV and other hepadnavirus genome sequences. The sequences for HBV genomes vary slightly in length from 3182 to 3221 base pairs. All of the sequences can be perfectly aligned, the variations in length being due to minor insertions or deletions as a result of point mutations particularly in the region immediately before the pre-S region (Fig. 3).

The short strand of the genome closed by the DNA polymerase is regarded as being positive in polarity, i.e. containing nucleotide sequences of the same sense as the corresponding mRNA. The complete long, or negative strand is linear and with a fixed length of approximately 3200 nucleotides, somewhat shorter than initial estimates from gel electrophoresis experiments. The circular nature of the genome is due to base pairing of the 5' ends of both strands extending for about 200 nucleotides. A low mol. wt. protein is covalently linked to the 5' end of the long, negative strand (Gerlich & Robinson, 1980), a feature that may be of importance in priming the synthesis of new positive strands from the negative strand template (see below).

The genetic organization of hepadnavirus genomes is similar. Analysis of the complete nucleotide sequence of all hepadnaviruses examined to date shows four open reading frames on the same large negative strand. These are termed S, P, X and C (Fig. 3, Table 3). Insertions or
deletions apparent from comparisons between these sequences show these conserving the reading frames delineated by start and stop codons. Assignment of the corresponding products has been obtained for at least two of the reading frames by comparison with the biochemical information on the viral proteins, notably the core and surface antigen polypeptides. Both these genes are read in different phases. These genes are overlapped by a major open reading frame in the same phase as the C gene which extends over almost the entire length of the genome and is believed to code for the polymerase protein, P. This overlap is very reminiscent of the situation in phage φX174 where sequencing studies have shown that the DNA genome may code for different polypeptide species by selective initiation at different points (Sanger et al., 1977). In the case of the hepadnavirus genome, the negative strand thus requires to be read one and a half times for full expression (Fig. 3). Interestingly, there is an apparent fusion between the C and X gene products for DHV, resulting in a core polypeptide almost twice the size of the core polypeptide of the mammalian hepadnaviruses (Newbold et al., 1984).

The S gene codes for the major 226 amino acid-long envelope polypeptide of mol. wt. 25 000 (p25) and this protein can also exist as a higher mol. wt. glycosylated form of mol. wt. 30 000 (gp30). Variable glycosylation of the major polypeptide can give rise to multiple bands on gel electrophoresis. Notably, p25 is not glycosylated in yeast-derived recombinant HBsAg (McAleeer et al., 1984). It is interesting to note that the extent of sequence homology in this region is reflected by the variation in antigenic cross-reactivity seen between the hepadnavirus HBsAg reactivities (Robinson et al., 1982). Transcription can be initiated from at least one, possibly two, initiation codons upstream from the S gene, in the so-called pre-S region. Minor components of HBsAg particles representing expression of these sequences may be resolved by SDS–PAGE in addition to the major gp30 and p25 proteins. These possess carboxyl termini in common with gp30 and p25 but have the additional amino-terminal sequences of either 55 or 175 amino acids in length and are termed pre-S2 and pre-S1 respectively (see Fig. 5). These three translation products appear unique, with no evidence of any precursor–product relationships. The appearance of pre-S sequences in HBsAg may be particularly dominant in infected individuals supporting high levels of virus replication (Stibbe & Gerlich, 1983). The nucleotide sequence in the pre-S region immediately adjacent to the S gene is highly conserved among all genome sequences so far examined, suggesting this may be a functionally important gene product. Indeed, a receptor for polymerized serum albumin has been located in this region; elimination of this product from HBsAg particles abolishes binding to polymerized albumin but expression is not required for assembly and secretion of the 22 nm spherical forms (Persing et al., 1985).

It has been suggested that the pre-S2 component of HBsAg may be important in the attachment of virus to the hepatocyte plasma membrane (Machida et al., 1983) and represent a dominant antibody-binding site (Milich et al., 1985). The inclusion of pre-S sequences into HBV vaccines in addition to the S gene products may help to overcome non-responsiveness to HBsAg seen in some vaccinees since the immune response to this region as measured in mice appears to be controlled by distinct major histocompatibility complex-linked genes. Considerably more variation is seen between sequences in the pre-S1 region: the use of the initiation codon at the beginning of this region would code for a 41 000 mol. wt. product, a polypeptide frequently seen in serum-derived or transfected cell cultures (Howard & Burrell, 1976; Laub et al., 1983).

Early cloning studies of restriction fragments showed the expression of HBCAg-positive proteins in E. coli (e.g. Burrell et al., 1979). This antigen is conformationally dependent as witnessed by the core-like structures seen under the electron microscope (Cohen & Richmond, 1982) and the change in antigenicity to an e antigen-positive state on treatment with ionic detergents (Mackay et al., 1981). The term C gene thus refers to the end product of translation and maturation rather than the primary gene product per se. There is some disagreement as to the exact site of the initiation codon for this protein, despite the fact that the majority of the sequence in this region of the genome is highly conserved. In contrast to the HBsAg polypeptides, the amino-terminal amino acid sequence of the naturally occurring molecule is unknown. The 22 000 mol. wt. protein reported by Feitelson et al. (1982) could be coded for by the sequences reported for HBV of the adw and ayw subtypes, but this is deleted in the other
sequences when a smaller reading frame utilizing the next initiation codon would give a smaller 19000 mol. wt. product. At the carboxyl-terminal end of the molecule, the sequence information suggests that two amino acids may be missing in the core polypeptide molecule expressed by HBV of subtypes other than adw. Sequencing of HBeAg-positive protein purified from serum suggests that this reactivity arises by proteolytic cleavage of 33 to 35 amino acids from the carboxyl terminus of the molecule to give a 15500 mol. wt. polypeptide (Takahashi et al., 1983). The significance of these differences is not yet clear.

The so-called X gene codes for a polypeptide equivalent to a 17000 mol. wt. protein 145 to 154 amino acids in length, according to the sequence examined. Although the function of this region is unknown, Moriarty et al. (1985) have identified the existence of this gene product in liver samples infected with HBV. Serological analysis indicates that a protein of 28000 mol. wt. contains cross-reactive epitopes with synthetic peptide structures made using the sequence data in this region, and that antibodies to this product are particularly prominent in individuals with primary liver cancer. This relationship has yet to be conclusively established, however, and the apparent larger size of the X product in vivo compared to the predicted molecule is unexplained.

The P gene is believed to code for the DNA polymerase, or at least a virus-specific functional subunit of this enzyme activity. However, the enzyme has yet to be purified free from virus cores in a functional state, thus severely hampering analysis of the native protein. However, Toh et al. (1983) have noted considerable homology between this reading frame and the pol gene of several oncogenic retroviruses and the polymerase of cauliflower mosaic virus. This homology appears to be highest in the S gene region. That this codes for a protein with the properties of a viral reverse transcriptase activity is especially attractive given the recent description of DHV replication occurring via an RNA intermediate (Summers & Mason, 1982; see next section).

**Replication and DNA synthesis**

Immunofluorescence analysis of infected liver clearly shows the presence of the HBcAg within the nuclei of infected hepatocytes (Barker et al., 1974). However, core components are also found in the cytoplasm closely associated with replicating complexes consisting of genome intermediates and morphologically discrete core particles. Gowans et al. (1984) have described varying levels of DNA expression in infected hepatocytes. Three different levels of genome expression were found in liver tissue from individuals persistently infected with HBV. Some cells containing low copy number of HBV genomes showed nuclear viral DNA in the absence of antigen expression, whereas others with low copy number showed either nuclear HBCAg or cytoplasmic HBsAg, or both. The latter cells were found particularly around the portal tracts and were predominant in anti-HBeAg-positive patients. Finally, scattered cells singly or in foci were found in HBeAg-positive patients with large amounts of cytoplasmic viral DNA and HBCAg.

Further information as to the nature of these cytoplasmic replicative intermediates has come from the study of Pekin duck liver infected with DHV where it has been estimated that the liver contains up to 100-fold more copies of the viral genome compared to the circulation, and immunofluorescence shows viral antigens in virtually all cells (Mason et al., 1984, 1985). Electron microscopy revealed these complexes as being of the larger diameter associated with the DHV core component, and were found capable of synthesizing both positive and negative strand copies of the DNA in vitro (Summers & Mason, 1982). Large amounts of free negative strands shorter than genome length were found in these complexes indicative of nascent DNA, but were not found attached to positive-strand DNA templates (Mason et al., 1982). The presence of strands of both polarities in these complexes was in marked contrast to the activity of the extracellular virion endogenous polymerase reaction which was found to synthesize positive-strand DNA only.

Positive-strand synthesis is almost totally inhibited by actinomycin D, an inhibitor of DNA-dependent RNA synthesis, whereas negative-strand synthesis continues, thus suggesting that negative DNA strands are synthesized on an RNA template of positive polarity generated in the complexes by reverse transcriptase-like activity from the parental negative DNA strand. These events in the replication of the hepadnavirus genome are outlined in Fig. 4. The size of the RNA
Fig. 4. Replication strategy of the hepadnavirus genome (see text). After the completion of the negative \((-\) strand by the endogenous polymerase (1) transcription is initiated. Full-length transcripts become inserted into maturing core particles later in the replication cycle (3) to form a template for reverse transcription (4). Removal of the template is then followed by synthesis of the positive \((+)\) DNA strand, which is not completed prior to virus maturation and release. A protein is attached to the 5' end of the \((-\) strand (boxed) which may play a role in priming \((+)\) strand synthesis.

component in the core complexes varies inversely with the size of the DNA component, possibly because the RNA positive-strand template is degraded by ribonuclease activity immediately it is transcribed into DNA. Synthesis of positive-strand DNA is therefore not detected in significant amounts in association with growing negative strands yet is sensitive to actinomycin D by virtue of requiring the previous step of RNA synthesis for the manufacture of a suitable negative-strand DNA template. Replicative intermediates containing growing positive DNA strands closely resemble the completed core component, with initiation of positive strands occurring close to the 5' end of the negative strand with elongation across the 3' end of the template generating the cohesive ends characteristic of the viral genome. That these events occur in close association with virus-specific C gene products suggests a close association between core protein maturation and packaging of newly synthesized DNA. Only full length negative DNA strands are encapsidated within immature cores, presumably preventing fresh RNA positive-strand synthesis. Further as yet undetermined packaging signals may also be involved in the final stage of complementary positive-strand DNA synthesis.

Many investigators have alluded to the possible generation of defective interfering (DI) particles playing a role in either the establishment or maintenance of persistent hepadnavirus infection. For example, Gerin et al. (1975) were able to separate HBV particles from serum into several fractions on the basis of DNA content. It is tempting to speculate, therefore, that in persistently infected cells, either smaller negative DNA strand templates become incorporated into immature core structures or positive DNA strand synthesis is prematurely terminated at the completion of virus maturation. This would give rise to the variable gap in the positive strands seen in extracellular virions of man and animals chronically infected with these viruses. In support of this hypothesis, Ruiz-Opazo et al. (1982) have shown that small amounts of viral DNA in the liver and serum of a chimpanzee persistently infected with HBV exist as complete double-stranded DNA structures which assume a supercoiled configuration. These structures may be 'relaxed' \textit{in vitro} by S1 nuclease activity directed at a hairpin loop structure within the unique site recognized by the restriction enzyme \textit{EcoRI}, and presumably a similar enzyme activity \textit{in vivo} may be required prior to primary transcription of the viral genome \textit{in vivo}. If correct, this would suggest that closure of the single-strand gaps on the positive DNA strand by the endogenous DNA polymerase is not required for initiating infection; in retrospect, the use of
Review: The hepadnaviruses

The finding that reverse transcription occurs during genome replication has prompted further review of genome organization in comparison with the retroviruses (Summers & Mason, 1982). Similarities include the unidirectional arrangement of the viral genes, the asymmetric synthesis of viral strands in the replicative complexes, and the presence of cohesive termini that may generate direct terminal repeats on the linear form of the genome (Sattler & Robinson, 1979). Priming of reverse transcription may differ, however. Unlike the requirement of cellular tRNA in the priming of cDNA synthesis from virion RNA, hepadnaviruses may utilize a protein covalently bound at the 5' end of the long, negative strand (Gerlich & Robinson, 1980; Molnar-Kimber et al., 1984). Although the size of this protein is unknown, Gerlich & Robinson (1980) estimated that up to 100000 mol. wt. of protein is attached per genome. This may represent either the product of the long open reading frame P (see Table 3) or a polymer of the newly characterized X protein (Moriarty et al., 1985). It has been noted by Mason et al. (1981) that infected duck liver contains a large amount of polyadenylated RNA almost the same size as the viral genome. Whether or not this represents the template for reverse transcription is not clear, although a certain amount becomes incorporated into viral cores. If this indeed represents the template, then the poly(A) tract at the 3' end of the RNA molecule would need to be removed early in the synthesis of the nascent negative DNA strand. Certainly this is the case in retroviruses where poly(A) is excised prior to, or during, the shift of enzyme activity from the 5' to the 3' end of the RNA template. The precise location of the RNA template relative to the hepadnavirus genome sequence also is unclear, although studies indicate the region of overlap contains two transcription sites which may be recognized by polymerase of host origin (Rall et al., 1983; Chakraborty et al., 1980). The site of the initiation for positive DNA strand synthesis has yet to be identified although the use of an RNA primer in a manner analogous to the retroviruses created by an RNAse H activity would seem possible.

Expression of gene sequences

The overwhelming mass of HBsAg-reactive material in the sera of HBV-infected individuals is present as small 22 nm lipoprotein particles. The mol. wt. of these structures is estimated to be $2.5 \times 10^6$, of which up to 30% may be accounted for by lipid (Howard & Burrell, 1976). A direct comparison between the S gene sequences of HBV and WHBV shows extensive homology between these two proteins both at the nucleotide and amino acid level (Galibert et al., 1982), consistent with the observed high degree of serological cross-reactivity between the woodchuck and human viral surface antigens (Millman et al., 1982). Similar comparisons with S gene sequences of the ground squirrel and duck viruses show less homology with the S gene of HBV (Mandart et al., 1984; Seeger et al., 1984). Indeed, the duck virus S gene contains an extensive deletion in that region of the sequence believed to code for the a group antigenic determinant of HBsAg between residues 110 and 150. The extent of cross-reactivity between the major antigens of each of the hepadnaviruses has been summarized by Robinson et al. (1982). Broadly speaking, WHBV HBsAg shows the closest antigenic relationship with the human virus antigen, with minor differences in amino acid composition being shared with some serotypes of HBV (Galibert et al., 1982). The surface antigen coded by GSHV is intermediate in antigenic relatedness between the woodchuck and duck hepadnaviruses.

The similarly sized 25000 (p25) and 30000 (gp30) mol. wt. components of human HBsAg 22 nm particles together form a subunit structure via disulphide bonds. Within the infected cell, a
proportion of the S gene product becomes glycosylated through the action of the host-specific glycosylation pathway. The role of carbohydrate side-chains is not known, however, although glycosylation is not essential for antigenicity as evidenced by the full retention of immunogenic capacity of HBsAg expressed in yeast in the absence of added sugar residues (McAleer et al., 1984). Of greater importance may be the presence of carbohydrate residues on the surface of intact 42 nm virions; Neurath et al. (1975) pointed out that terminal sialyl residues may determine the half-life of HBsAg in the circulation. Highly charged residues on the surface of the virus may therefore be important for binding to the appropriate cellular receptors of susceptible cells. Further information on the biochemical properties of HBsAg expressed by human HBV may be found in Howard (1981).

At first sight, expression of HBsAg polypeptide from the viral genome appears to occur by transcription of the S gene, accounting for approximately 21% of the total genetic capacity of the virus. The close agreement between the length of this sequence and physicochemical data on the HBsAg polypeptides would appear to eliminate the need for mRNA splicing if indeed transcription was restricted to this open reading frame. However, an examination of the sequence of the whole genome together with analysis of mRNA transcripts produced both in hepatoma cell lines and cells transfected with cloned sequences suggest that the hepadnavirus genome utilizes an unusual and complex strategy of genome transcription. Two major polyadenylated transcripts have been characterized (Fig. 3). The first, small mRNA is 2-1 kilobases long and covers the S and X genes in the same phase, ending at the beginning of the C gene. The second, larger 3-5 kilobase mRNA represents transcription of the C and P genes. The size of this larger transcript implies that covalently closed DNA is required. Thus, full expression of the hepatadnavirus genome requires two cycles of transcription.

Two lines of evidence suggest that the promoter for the S gene lies upstream beyond the pre-S region and that the 25,000 mol. wt. S gene product is generated by post-translational cleavage. Firstly, Rall et al. (1983) have shown that in vitro transcription of truncated fragments of the genome by RNA polymerase II requires the presence of this promoter region. Secondly, HBsAg production in transfected L cells is optimal when the cloned sequence introduced into the fibroblast includes the same sequence (Pourcel et al., 1982). However, HBsAg may still be expressed in the cells even if the pre-S region is deleted (Valenzuela et al., 1982; Laub et al., 1983; Crowley et al., 1983) and that initiation of transcription may occur at the beginning of the S gene is indicated by a methionine residue at the amino-terminal residue of the 25,000 mol. wt. HBsAg polypeptide.

The availability of cloned fragments has allowed further analysis of transcription signals. This has shown that a sequence within the pre-S region may be of greater importance in the initiation of the S gene transcripts (Cattaneo et al., 1983). At approximately 185 nucleotides upstream of the start of the S gene there is extensive sequence homology in the pre-S region between HBV and the promoter of the major late transcript of simian virus 40 (SV40). Cattaneo and colleagues studied RNA extracts prepared from infected chimpanzee livers as well as transcripts from transfected cell cultures, thereby validating these experiments for analysing events accompanying HBV infection in vivo. These findings have been complemented by the work of Standring et al. (1984) who found that the putative promoter preceding the pre-S region is not essential either for HBsAg expression or for the assembly and maturation of HBsAg particles. Interestingly, this study also showed that there are few pre-S region transcripts in HBsAg mRNA from transfected cells although the size of the mRNA (2-3 kilobases) would suggest transcription from a site further upstream. This places the 3' end of such transcripts as extending well beyond the S gene sequence within the open reading frame coding for the core polypeptide (C gene). As the viral genome within this region contains a sequence for signalling the addition of a poly(A) tail to the transcript which spans the S gene (Cattaneo et al., 1983), this has implications for understanding the expression of the C gene which presumably can occur only by a transcription mechanism which ignores this sequence. These events are outlined in Fig. 5. The exact 5' end of the mRNA coding for HBsAg may be heterogeneous within the pre-S region, but only the largest transcripts allow the translation of both the major 25,000 mol. wt. (p25) HBsAg polypeptide and the additional 55 amino acids of the pre-S2 region known to contain the binding
site for polymerized serum albumin (Machida et al., 1983). Whether the p25 molecule is produced by the cleavage of the larger 281 amino acid polypeptide containing both S and pre-S2 sequences or whether it is the product of internal initiation of translation in the 2.1 kilobase mRNA is not known. Several workers have speculated that the sequence diversity between hepadnaviruses in the pre-S region may determine species specificity, particularly in the larger pre-S1 region towards the promoter recognized by RNA polymerase II. Rutter et al. (1984) estimate that approximately 10% of mRNA transcripts in hepatoma cells coding for HBsAg cover the entire pre-S region: thus, limited expression of these extra sequences may be important in determining viral tropism.

The expression of the hepadnavirus core proteins (Fig. 6) has proved more difficult to analyse partly because of the difficulty in either separating a sufficient quantity of HBcAg from virus particles free of HBsAg or partly because of the formation of immune complexes during the process of extracting core antigen from infected livers. As a result, there are several conflicting reports as to the physicochemical nature of the core particles from these sources (for review, see Howard, 1981). There is general agreement, however, that the glycoproteins are absent, although phosphorylation, of unknown function, of HBV core particles may occur (Albin & Robinson, 1980). Most studies have also shown the existence of a major polypeptide in the mol. wt. range of 16000 to 20000, which is in good agreement with the predicted size of about 19000 for the C gene product. Feitelson et al. (1982) have found a 22000 mol. wt. HBcAg-reactive polypeptide in infected liver which probably undergoes cleavage to give rise to a 19000 mol. wt. product; Cohen & Richmond (1982) have found that the HBcAg expressed in E. coli spontaneously assembles into core-like particles morphologically indistinguishable from native structures, illustrating that the basic information required for core assembly is encoded by the C gene. Given the role of maturing core particles within genome synthesis, a series of conformational changes may be expected as virus maturation proceeds. The involvement of other virus-specific (or host specific) proteins in minor amounts cannot as yet be determined, although higher mol. wt. components are often seen in liver-derived core preparations (Hruska & Robinson, 1977; Fields et al., 1977). Neurath et al. (1978) found that core antigen prepared

Fig. 5. Expression of S gene products. (a) Transcription is controlled by several promoters, all of which give rise to mRNA extending beyond the end of the S gene, possibly into the region controlling the expression of the C gene. (b) Translation of the various mRNA species produces a number of polypeptides, all of which have the common S polypeptide of approximately 25000 mol. wt. Larger translation products have the additional pre-S sequences, many of which may be glycosylated. Map positions on the genome are those of Valenzuela et al. (1979).
from sera with or without HBeAg was heterogeneous with respect to isoelectric point; core antigen from HBeAg-positive serum produced two fractions separated by centrifugation in contrast to only one from HBeAg-negative serum. This suggests that HBeAg-positive cores may exist in different conformations independent of whether or not DNA is present. The presence of arginine-rich regions towards the carboxyl end of the 21000 mol. wt. predicted C gene product is compatible with the binding of core polypeptide molecules to nucleic acid. HBeAg activity is derived from the amino-terminal portion of the core polypeptide; these determinants may be expressed by either conformational change in the presence of detergents or result from proteolytic cleavage of HBeAg-positive protein (Yoshizawa et al., 1979; Takahashi et al., 1979; Mackay et al., 1981). In this context, it is interesting to note that Ohori et al. (1979) found that the release of HBeAg from disrupted virions was greatest for preparations containing both DNA polymerase and the DNA genome; this suggests that polypeptides bearing HBeAg determinants may have a function during virus maturation and assembly.

Rutter et al. (1984) have pointed out that C gene transcription starts close to those regions of the genome centrally involved in the regulation of expression of the entire genome. Tandem expression of the HBV genome with cellular sequences has been achieved in transfected cells using a multicopy sequence of the genome (Gough & Murray, 1982). Two RNA transcripts transformed cell lines producing both HBcAg and HBeAg specificities. These were absent in those lines producing HBsAg only. This would indicate that C gene expression occurs by transcription proceeding through polyadenylation/termination signals in a manner analogous to the late transcripts encoded by polyoma virus (Tooze, 1981). It is always possible, however, that
study of transfected cell cultures does not faithfully represent the situation in naturally infected cells; splicing may occur within the infected hepatocyte unrelated to translation of the C gene sequence but necessary for full virus replication. The close association of HBCAg and the nuclei of infected hepatocytes has also yet to be explained; however, it is conceivable that transcription of the genome may take place in the nucleus, particularly if certain transcripts require host-dependent splicing whereas replication of the genome may occur predominantly in the cytoplasm. Current interest in the possible expression of HBCAg activity at the plasma membrane of infected hepatocytes with implications for understanding immunopathology and immunity (Iwarson et al., 1985; Mondelli et al., 1982) would require the C gene protein to interact in some way with integral membrane proteins. An examination of the predicted HBCAg polypeptide amino acid sequence has not revealed a hydrophobic domain, thus opening the possibility that the translation product of the C gene may associate either with a host protein, a hitherto unrecognized HBV-specific product, or is linked to the hydrophobic region of the S gene product by splicing of the long transcript covering the entire genome.

Interest in the C gene transcription centres around the possible co-regulation of all genome transcription and the close proximity of the promoter to putative sites of integration into cellular DNA. This region contains many of the features found in the long terminal repeats (LTRs) of retroviruses, including promoters for transcription, direct and inverted repeat sequences and tracts rich in purines. Comparison with the WHBV sequence in this region shows that both the purine-rich tracts and the direct repeat sequences have been conserved (Galibert et al., 1982). Rutter et al. (1984) have proposed that the purine-rich tract on the positive strand may be involved in priming DNA synthesis during genome replication and the direct repeat sequences on the negative strand are directly involved in the formation of the nicks by directing either the initiation or trimming of newly formed minus DNA strands during the process of reverse transcription.

**Hepatitis B virus and hepatocellular carcinoma**

The close association between HBV and human primary liver cancer has been well documented (Zuckerman, 1982). Considerable advances have been made in understanding genome expression in hepatoma cells either maintained in culture or by direct examination of liver tissue obtained by biopsy. The cardinal finding is that long-term infection with HBV is associated with integration of virus-specific sequences into cellular DNA (Brechot et al., 1981, 1982; Shafritz et al., 1981; Chen et al., 1982; Kam et al., 1982). Integration appears to be clonal in human livers with focal regeneration occurring subsequent to integration, and has also been described in infected woodchucks (Dejean et al., 1982). To date, integration has not been described for the duck or ground squirrel viruses and, as already discussed, is not essential for virus replication.

The PLC/PRF/5 hepatoma cell line, derived by J. Alexander and colleagues (MacNab et al., 1976) has been intensively studied. These cells produce 22 nm particles typical of HBsAg found in human serum but do not express HBCAg or produce infectious virus (Copeland et al., 1980; Tabor et al., 1981). Although some four to six genome equivalents are distributed at up to six integration sites in host chromosomal DNA, hybridization studies have clearly demonstrated the absence of full-length intact HBV genomes in favour of extensive fragmentations, insertions and deletions (for review, see Rutter et al., 1984). Other hepatoma cell lines have been studied, notably the HEP-3B line of Aden et al. (1979). This contains integrated DNA at two discrete sites and one subline was found to contain a complete genome sequence (Twist et al., 1981). Although one clone of the PLC/PRF/5 cell line has been found to contain the complete C gene (Rutter et al., 1984), extensive methylation of the nucleotide sequence may inhibit transcription, thereby accounting for lack of HBCAg expression in these cells (Miller & Robinson, 1983). This is in accord with the absence of detectable transcripts spanning the C gene sequence (Edman et al., 1981).

Parallel studies using hepatoma tissue from HBV carriers in various areas of the world have shown integration of virus-specific DNA in almost every tumour where the presence of the HBV genome has been demonstrated (for review, see Zuckerman, 1985). As with the hepatoma cell
line, integration appears to occur at random, although bands of viral DNA of equivalent size may be demonstrated between tumours (Shafritz & Kew, 1981). Similar evidence of integration is available for WHBV (Ogston et al., 1982). Instances where both integrated and episomal viral DNA have been demonstrated have been suggested as representing the early stages of tumour development (Brechot et al., 1981). However, the degree of involvement of free genetic material in carcinogenesis is difficult to assess, especially as DNA extracts may become contaminated with virus from blood. There are a number of reports suggesting that healthy liver tissue adjacent to tumours shows both integrated and non-integrated DNA sequences (Gerin et al., 1982; Shafritz et al., 1981; Brechot et al., 1981). Despite the variation in integration sites between samples, the finding of a unique pattern in each individual points to the clonal nature of HBV-associated tumours whereby a single progenitor cell has been stimulated to divide by the transformation event. The approximate relationship between viral genome copy number per cell and the number of integration sites in hepatoma tissue also suggests the monoclonal nature of tumours and that multiple sites of integration arise by DNA rearrangements of the original integration (Robinson et al., 1984).

There is considerable interest in analysing the virus–host DNA junction sequences in order to ascertain the mechanism of genome integration. At the present time, there is no evidence to suggest that the flanking cellular sequences represent known oncogene sequences or that such sequences are affected by genome integration (Fung et al., 1984). This is in clear contrast to the proposed promoter-insertion mechanism associated with retrovirus-induced transformation whereby viral LTR sequences appear to direct the expression of cellular oncogenes (Varmus & Swanstrom, 1982). Some rearrangement of cellular sequences may occur on integration, however. Shafritz & Rogler (1984) have found that whereas the restriction endonuclease sites to either right or left of HBV sequences integrated into hepatocellular tissue correspond to normal cellular DNA, these host sequences could not have been contiguous prior to insertion. These studies showed that up to 12 kilobases of host DNA surrounding the integration site were modified, either by deletion or by translation of sequences from other positions in the host genome. The possibility arises, therefore, that HBV integration may stimulate a rearrangement of cellular oncogenes, as has been demonstrated to occur in Burkitt's lymphoma (Yunis, 1983). The junctions of HBV sequences in the PLC/PRF/5 cell line are distributed over the entire genome, although the study of Rutter et al. (1984) found a distinct preference for the region between the C and S genes. Virus–virus junctions were more restricted, many occurring immediately before or in the pre-S coding region. Again, this is in contrast to the retroviruses where there is a single integration site and the proviral state of the genome is maintained. Koshy et al. (1983) studying human tissue and Shafritz's laboratory (Ogston et al., 1982; Shafritz & Rogler, 1984) using the woodchuck model have slightly different evidence to suggest that integration occurs predominantly in the cohesive end region or in the single-stranded region as characterized within the virus particle. Interaction of the single-stranded region with cellular DNA during cell growth would then lead to incorporation of viral sequences by a process of recombination. Interestingly, the integrated woodchuck sequences often encompass the X gene region; as mentioned previously, Moriarty et al. (1985) have recently identified antibody to a protein coded by the HBV X gene sequence in patients with hepatoma, although point out that further work is required to examine this relationship. An antigenic reactivity specific for the nuclei of the two human hepatoma cell lines PLC/PRF/5 and HEP-3B has been described by Wen et al. (1983), although the identity and function of this new specificity have yet to be determined.

Few tumour cells have been found to express viral antigens in sufficient quantities to allow detection by serological methods. HBsAg production can be detected by immunofluorescence in PLC/PRF/5 cells (Wen et al., 1981) but hepatoma material from HBsAg-positive primary liver cancer patients is invariably negative (Kew, 1978). HBeAg is found even more rarely, probably as integrated sequences tend to undergo extensive rearrangements in the promoter region for C gene expression. The evidence discussed above indicating that complete genome transcripts are required for HBeAg expression together with the methylation of C gene sequences also mitigate against HBeAg production in cells containing integrated hepadnavirus sequences. Also,
integrated DNA is usually present in much smaller amounts, often less than one copy per cell, compared to cells with free viral DNA where active virus transcription may be occurring from over 500 copies of the viral genome per cell (Kam et al., 1982). Transcripts from both the pre-S and S regions of the genome are clearly seen by sensitive hybridization techniques (Edman et al., 1981). These account for about 0.05% of the total cellular mRNA content of these cells (Rutter et al., 1984), representing transcription of virus-specific pre-S sequences at the 5' end through the S gene to be variably terminated at stop signals in the cellular flanking sequences to the right of the integrated sequence. Interestingly, these mRNAs include a viral sequence downstream to the S gene which behaves as an enhancer of S gene expression independent of the gene promoter. Such enhancer sequences have been described first in SV40 and may act at the 3' end of the stimulated gene. In preliminary experiments, Rutter et al. (1984) have shown that enhancement is particularly strong in liver cells.

Integrated HBV DNA has been detected in human carriers persistently infected for at least 2 years (Shafritz et al., 1981; Kam et al., 1982) although integration may occur even earlier (Brechot et al., 1981). Non-integrated high molecular weight viral DNA has also been found in ground squirrels and woodchucks persistently infected with their respective viruses (Rogler & Summers, 1982; Marion et al., 1980). Many human HBV carriers with integrated sequences show no histological signs of liver disease and are anti-HBe-positive, suggestive of low or negligible virus production. The demonstration of integrated sequences in these patients (Shafritz & Arias, 1982) has clear implications for the treatment of the carrier state in such cases where antiviral therapy or immunological stimulation or both would be ineffectual. The numbers of carriers may be supplemented by individuals who are not expressing the HBsAg marker of HBV carriage but contain integrated sequences. Brechot et al. (1981) have shown that individuals positive for anti-HBs may harbour at least some HBV-specific sequences and Brown et al. (1983) showed that low quantities of HBsAg continue to be produced in individuals with primary liver cancer positive for anti-HBs. Integration of the genome, successive rearrangements of viral gene sequences promoted by environmental or physiological factors and ultimately the induction of the neoplastic state are thus all consequences of the host's failure to limit virus activity during acute infection at any one of a number of stages in the complex replication cycle of these viruses.

**REFERENCES**


Review: The hepadnaviruses


