Biology of Hepatitis B Virus

Pierre Tiollais, Patrick Charnay, Girish N. Vyas

Summary. Immunochemical investigations of the viral antigens and molecular characterization of the viral DNA have elucidated the nature of the hepatitis B virus infection underlying acute, chronic, and oncogenic disorders of the liver in man. Cloning and sequencing of viral DNA have made possible studies on the structure of the genome and on certain aspects of the biology of the virus, hitherto constrained for a lack of tissue culture systems and laboratory animal models useful in its propagation.

The persistence of hepatitis B virus (HBV) in an estimated 200 million persons in the world, the frequent occurence of occult infection in close contacts of the carriers, perinatal vertical transmission from mother to infant, and high incidence of insidious chronic liver disease or cancer of the liver among the carriers signify a major public health problem of worldwide concern (*1–4*). Understanding the biology of the virus and defining the immunogenetic mechanisms of host responses are, therefore, crucial to the development of safe and practical procedures for control of HBV infection. A contemporary assessment of etiology, epidemiology, pathogenesis, and prevention of viral hepatitis has been made (*4*). In this article we present the current concepts and recent biologic information on antigens and genetic organization of the HBV.

Viral Particles

Immunologic markers of HBV infection include the surface antigen (HBsAg), the core antigen (HBcAg), the "e" antigen (HBeAg), and their respective antibodies. Infected plasma contains varying amounts of particles of different sizes and forms (*2*). The common spherical and filamentous particles with a mean diameter of 22 nanometers are devoid of DNA and represent free envelopes of the virus. The less common 42-nanometer Dane particles represent the virion and consist of an envelope and a 27-nanometer nucleocapsid that contains a molecule of DNA (*5, 6*). Free nucleocapsids are observed in the nucleus of infected hepatocytes but not in the plasma.

The infected hepatocytes synthesize an excessive amount of envelope, which circulates with a half-life of 3.3 days (*7*). Because antibodies against HBsAg are protective against HBV infection, the virus-free envelopes present in the plasma of chronic carriers can effectively be used as a source of vaccine (*8*). Therefore, these particles have been a subject of extensive immunochemical characterization with regard to the HBsAg.

Viral Envelope and Surface Antigen

The HBsAg in viral envelope has one well-characterized group-specific determinant a and two sets of mutually exclusive subtype determinants d/y and w/r. Thus four major subtypes of HBsAg—adw, ayw, adr, and ayr—denote the phenotypes of the virion (*9*). Subdivision of a specificity into a_1, a_2, a_3, and other intermediate specificities has led to serologic classification of HBsAg into ten known categories (*10*).

The viral envelope contains typical membrane lipids and proteins, some of which are glycosylated. There is no unanimity about the number and precise size of the constituent polypeptides; however, two major polypeptides termed PI and PII are present in about equal quantity. The third major polypeptide of about 70,000 daltons cross-reacts with albumin (*11*). Other large-molecular-weight polypeptides are considered aggregates of PI and PII (*11*). Similarity in amino acid composition of PI and PII and identity of NH_2-terminal and COOH-terminal sequences suggest that PI and PII may be identical, except that PII is glycosylated (*12*). This relation was confirmed by identification in HBV DNA of a single nucleotide sequence (defined below as the gene S) which codes for a 25,400-dalton polypeptide consisting of 226 amino acid residues (*13–17*). The three recorded sequences of PI are compared in Fig. 1A. Theoretical analysis of the primary structure suggests that there are three regions of β sheets (residues 1 to 31, 75 to 109, and 157 to 226) and two regions of β turns with little or no helical structure (*18*). The two hydrophilic regions (residues 32 to 74, and 110 to 156) defined by β turns are separated by a hydrophobic stretch 19 amino acids long (Fig. 1). This stretch may serve as a signal sequence for insertion of PI in the envelope or permit hydrophobic interactions between molecules of PI or its derivatives. The central region (residues 110 to 156) of PI is particularly hydrophilic and essentially it contains the lysine-dependent HBsAg a determinants (*18*). Eight variations between residues 114 to 143 in the three known sequences make this hydrophilic region a candidate for the heterogeneity of the antigenic determinants of HBsAg and their experimental analogs which may prove useful as synthetic antigens (*10, 18*).

The sequence Asn-X-Thr (or Ser) is observed thrice and could potentially serve as the site for attachment of *N*-glycosidic residues. However, the first one (position 3 to 5) is in a hydrophobic region and is less likely to be a site. Since 75.8 micrograms of carbohydrate per milligram of total protein of the envelope has been reported (*19*), the actual molecular size of PII can be estimated to be around 29,000 daltons. The role of the carbohydrate moeity in antigenic activity of HBsAg has been controversial (*20*). The loss of HBsAg activity by periodate treatment may simply be due to alterations of protein conformation (*21*). For a full expression of the HBsAg activity, the interaction between PI and PII appears to be critical. Skelly, Howard, and Zuckerman used Triton X-100 to dissociate the envelope protein into two frac-

Dr. Tiollais is a professor in the Faculty of Medicine Lariboisière Saint-Louis, University of Paris VII, and director of the Unité de Recombinaison et Expression Génétique, Institut de la Santé et de la Recherche Médicale, INSERM U 163, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France. Dr. Charnay is an attaché de Recherche at INSERM and a staff scientist of the Unité de Recombinaison et Expression Génétique. Dr. Vyas is a professor of laboratory medicine and director of Blood Bank, University of California, San Francisco 94143.

0036-8075/81/0724-0406$01.50/0

tions, namely, albumin and a dimer composed of PI and PII (*22*). Mishiro *et al.* have further shown that both the antigenicity and immunogenicity of the dimer are comparable to the intact envelope and are reduced by more than 90 percent after treatment with β-mercaptoethanol (*23*). This is consistent with the earlier observation that reduction and alkylation of the envelope protein causes loss of antigenicity and immunogenicity due to conformational changes (*24*). The organization of the structural polypeptides PI and PII must be highly ordered by hydrophobic interactions and disulfide bonds whose precise intra- and interchain linkages are unknown at present. A model of the viral envelope is shown in Fig. 2.

The Nucleocapsid and Its Antigens

The nucleocapsid carrying the core antigen (HBcAg) is predominantly composed of a major polypeptide of 19,000 daltons (P19) (*25*). The amino acid sequence of this polypeptide (Fig. 1B) has been deduced from the nucleotide sequence defined below as the gene C (*16*). The polypeptide of 183 to 185 amino acids is composed of two parts. The first 140 residues of the amino terminal portion show nothing remarkable, but the 40 residues of the COOH-terminus manifest a protamine-like structure with repeating arginines separated by serine and proline (*26*). This portion of the molecule is probably involved in binding to the DNA molecule. Albin and Robinson have reported a protein kinase activity associated with the core of the Dane particles and shown that P19 is phosphorylated by the kinase (*27*).

The invariable occurrence of antibody to HBcAg in the plasma of chronic carriers does not permit detection of free cores of HBV in the circulation. The HBcAg is revealed by removal of the envelope after treatment of the circulating virion with the nonionic detergent (NP40) (*2*). Both HBcAg and HBeAg have been localized with the use of fluorescent antibodies in the nuclei of infected hepatocytes. However, only HBeAg is detectable in the plasma. This may be explained by the evidence that purified cores, when disrupted by treatment with sodium dodecyl sulfate, liberated P19, an experiment that showed both the antigenicity and immunogenicity of HBeAg (*28*). The association of HBeAg with various plasma proteins and the antigenic heterogeneity of HBeAg (e_1, e_2, and e_3) remain unclear (*29*, *30*).

A

```
1   Met Glu Asn Ile Thr Ser Gly Phe Leu Gly Pro Leu Leu Val Leu Gln Ala Gly Phe Phe

21  Leu Leu Thr Arg Ile Leu Thr Ile Pro Gln Ser Leu Asp Ser Trp Trp Thr Ser Leu Asn
                    Ser Pro                                             Thr
41  Phe Leu Gly Gly Thr Thr Val Cys Leu Gly Gln Asn Ser Gln Ser Pro Thr Ser Asn His
                    Thr Thr                                             Ile
                                 Ile
61  Ser Pro Thr Ser Cys Pro Pro Thr Cys Pro Gly Tyr Arg Trp Met Cys Leu Arg Arg Phe
                                 Thr
81  Ile Ile Phe Leu Phe Ile Leu Leu Leu Cys Leu Ile Phe Leu Leu Val Leu Leu Asp Tyr
                                                    Thr                         Pro
101 Gln Gly Met Leu Pro Val Cys Pro Leu Ile Pro Gly Ser Ser Thr Thr Ser Thr Gly Pro
                                                    Ser                         Ser
        Lys         Thr     Pro             Asn         Phe
121 Cys Arg Thr Cys Met Thr Thr Ala Gln Gly Thr Ser Met Tyr Pro Ser Cys Cys Csy Thr
        Arg         Thr     Pro             Ile         Tyr
            Thr                                                         Ala
141 Lys Pro Ser Asp Gly Asn Cys Thr Cys Ile Pro Ile Pro Ser Ser Trp Ala Phe Gly Lys
            Ser                                                         Gly
    Tyr                     Val
161 Phe Leu Trp Glu Trp Ala Ser Ala Arg Phe Ser Trp Leu Ser Leu Leu Val Pro Phe Val
    Phe                     Ala
                                Thr                 Ala
181 Gln Trp Phe Val Gly Leu Ser Pro Thr Val Trp Leu Ser Val Ile Trp Met Met Trp Tyr
                                    Ile             Val
                                    Val             Ile
201 Trp Gly Pro Ser Leu Tyr Ser Ile Leu Ser Pro Phe Leu Pro Leu Leu Pro Ile Phe Phe
                                    Leu             Leu
                Val
221 Cys Leu Trp Val Tyr Ile
                Ala
```

B

```
1   Met Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ala Thr Val Glu Leu Leu Ser Phe Leu Pro

                                                    Ser                 Glu
21  Ser Asp Phe Phe Pro Ser Val Arg Asp Leu Leu Asp Thr Ala Ser Ala Leu Tyr Arg Glu
                                                    Ala                 Asp

41  Ala Leu Glu Ser Pro Glu His Cys Ser Pro His His Thr Ala Leu Arg Gln Ala Ile Leu

                Glu                                 Asn
61  Cys Trp Gly Glu Leu Met Thr Leu Ala Thr Trp Val Gly Val Asn Leu Glu Asp Pro Ala
                Asp                                 Thr
                            Asn                     Val             Ile
81  Ser Arg Asp Leu Val Val Ser Tyr Val Asn Thr Asn Met Gly Leu Lys Phe Arg Gln Leu
                            Ser                     Val             Phe
                                                                    Leu
101 Leu Trp Phe His Ile Ser Cys Leu Thr Phe Gly Arg Glu Thr Val Ile Glu Tyr Leu Val
                                                                    Leu

121 Ser Phe Gly Val Trp Ile Arg Thr Pro Pro Ala Tyr Arg Pro Pro Asn Ala Pro Ile Leu

                                            Asp Arg
141 Ser Thr Leu Pro Glu Thr Thr Val Val Arg Arg Arg Gly Arg Ser Pro Arg Arg Arg Thr

                                    Pro
161 Pro Ser Pro Arg Arg Arg Arg Ser Gln Ser Pro Arg Arg Arg Arg Ser Gln Ser Arg Glu
                                    Gln

181 Ser Gln Cys
```

Fig. 1. Amino acid sequences of the major envelope protein shown in (A) and the major capsid protein shown in (B). The sequences were derived from the nucleotide sequence by Charnay *et al.* (*14*) and Galibert *et al.* (*15*). The amino acid changes observed in the sequences by Valenzuela *et al.* (*17*) are denoted above and those in the sequences by Pasek *et al.* (*16*) are denoted below the respective residues. The numbers in the left-hand columns indicate the serial numbers of the residues shown in rows of 20 amino acids. In (A) the conserved stretch of 19 amino acid hydrophobic sequence and the two boxed sequences of Asn-X-Thr (Ser) in hydrophilic regions potentially useful for carbohydrate attachment in the polypeptide PII. In (B), the protamine-like sequence, nine amino acids long, repeated twice at the COOH-terminal portion may serve the function of attachment to the encapsidated viral DNA. The two additional amino acids, observed in the sequence by Valenzuela *et al.* (*17*), are inserted. The abbreviations for the amino acid residues are Ala, alanine; Arg, arginine; Asp, aspartic acid; Asn, asparagine; Cys, cysteine; Gly, glycine; Gln, glutamine; Glu, glutamic acid; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Val, valine; X, any.

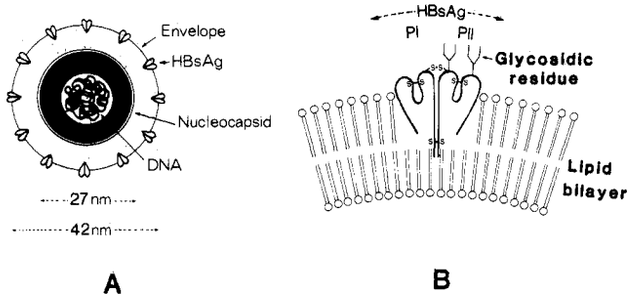

Fig. 2. Model of the viral envelope. (A) Model of the viral envelope showing the lipid bilayer with a dimeric subunit formed by polypeptide PI and glycopeptide PII, held together by a combination of inter- and intrachain disulfide bonds conferring on the protein a conformation essential for the full expression of the antigenicity and the immunogenicity of HBsAg. (B) Two hydrophilic regions are externalized in the model to represent areas containing various antigenic determinants. In PII, the one or two potential sites for attachment of the carbohydrate moiety are also depicted to reflect posttranslational glycosylation in the hydrophilic regions.

A new system of δ antigen (δAg) and its specific antibody (anti-δ) associated with HBV infection has been defined by nuclear immunofluorescence of infected hepatocytes (*31*). Although δAg is not detectable in the serum, it has been purified from nuclei of infected hepatocytes and characterized as a 68,000-dalton protein (*31*). The biology of both δAg and anti-δ is analogous to that of HBcAg and its antibody. It seems that δAg is included in HBsAg particles, and an RNA molecule has been reported to copurify with δAg. Transmission studies in chimpanzees suggested that δAg is associated with an agent requiring the functions of HBV for its expression and replication. This led Rizetto, Purcell, and Gerin to consider the δ agent as a defective virus and to suggest that the encapsidation of this agent with HBsAg could allow it access to the hepatocytes (*32*).

Physical Structure of the Viral Genome

The presence of a circular DNA molecule and an endogenous DNA polymerase activity in the nucleocapsid of the Dane particles has been well established (*33*); the genome, about 3200 nucleotides long, has a characteristic structure with a single-stranded region of variable length. Summers *et al.* proposed a model of two unequal strands, one long strand (L strand) of fixed length and a second short strand (S strand) of variable length (*34*). Recent experiments have validated and refined this model (*35–39*) (Fig. 3). The single Eco RI cleavage site is used as the origin of the physical map (*34–36*). However, such an Eco RI site is absent in certain HBV genomes (*37*). The L strand has a nick or a gap of a few nucleotides located at a fixed position around nucleotide 1800 (*35, 36, 39*), perhaps at nucleotide 1818, according to the results of Pasek *et al.* (*16*). The 5′ extremity of the S strand is located at a fixed position around nucleotide 1560, but the position of the 3′ extremity in different molecules is variable (*35, 36, 39*). The single-stranded region represents 0 to 50 percent of the molecule. The maintenance of the circular structure is assured by base-pairing of the 5′ extremities of the two strands over a length of 250 to 300 nucleotides (*39*). A covalently linked protein has been found at the 5′ extremity of the L strand, but its biological role is undefined (*40*). The DNA polymerase associated with the viral capsid is capable of repairing the gap by elongating the 3′ end of the S strand (*34*). The possible biological function of the characteristic structure of the DNA molecule has not been clearly defined. No remarkable sequence was revealed around the nick of the 5′ extremity of the S strand. The nick could be involved in the integration of HBV DNA in the infected cells.

Several restriction maps of the genome from different HBV subtypes reveal many differences (*13, 16, 35, 36*). Siddiqui *et al.* have shown that within the same subtype the restriction patterns are generally identical with occasional differences (*36*). This is not surprising since the restriction sites reflect the nucleotide sequence of the entire DNA molecule and the antigenic subtypes correspond only to a small portion of the genome. Sninsky *et al.* have shown no difference in the restriction pattern of different clones derived from the same serum (*38*). Therefore, it would appear that within a given chronic carrier the HBV genome is generally homogeneous.

Genetic Organization

Determination of the complete nucleotide sequence of the cloned DNA has provided definitive evidence that HBsAg and HBcAg are encoded by the DNA of Dane particles and that both genes are present on the same DNA molecule. A comparison of the coding capacity of the viral genome and the size of the viral polypeptides led us to conclude that the genome could code for all the known HBV proteins.

Among the three nucleotide sequences recorded, two cover the entire genome of 3182 and 3221 nucleotides reported by Galibert *et al.* (*15*) and Valenzuela *et al.* (*17*), respectively. When we omitted the two additional sequences of 6 and 33 nucleotides in the sequence reported by Valenzuela *et al.* (*17*), the three sequences were perfectly aligned for comparison without any other insertions or deletions. A 2 by 2 analysis of the three nucleotide sequences (*15–17*) demonstrated a relatively high degree of divergence. The nucleotide differences were 9.5 percent between Galibert *et al.* (*15*) and Valenzuela *et al.* (*17*), 4.1 percent between Galibert *et al.* (*15*) and Pasek *et al.* (*16*), and 11.2 percent between Pasek *et al.* (*16*) and Valenzuela *et al.* (*17*). No region contains a marked divergence or a mutational hot spot. Because of inadequate information on the viral proteins and almost none on the viral mRNA's (messenger RNA) and their processing, it may be premature to analyze the genetic organization solely from the nucleotide sequence. Nevertheless, transcription of nucleotide sequence and search for stop codons in the three reading frames permit localization of large open regions (without stop codons) potentially coding for viral proteins. While the L strand transcript contains four large potentially coding regions exactly aligned in the three sequences, the S strand transcript contains only short open regions whose positions are not conserved. Moreover, the translational amino acid sequences of the four open regions of L strand are better conserved than those derived from reading of the genome in frames considered noncoding (Table 1). These observations suggest that the open regions of the L strand code for proteins, whereas the S strand does not code for any protein.

The four open regions of the L strand are termed S, C, P, and X (Fig. 3). The region S can be divided into two parts, gene S and pre-S. Gene S, defined by nucleotides 155 to 883 and coding for envelope protein PI, has been localized on the genome by matching the amino and carboxyl terminal amino acid sequences of PI with the corresponding

nucleotide sequence of cloned HBV DNA (*12–14*). The methionine codon at position 155 is not the first one in the region S. The gene S is preceded by a pre-S sequence between nucleotides 2848 to 154 with a coding capacity for 163 amino acids (Fig. 3). The amino acid sequence of pre-S is less conserved than that of the sequence of polypeptide PI (Table 1). Nevertheless, the identical location and length of the pre-S sequence in the three sequences suggest a possible coding function for pre-S sequence. Actually, we cannot exclude at present the possibility that the polypeptide PI is cleaved from a precursor protein corresponding to the entire region S. The 33 additional nucleotides found by Valenzuela *et al.* (*17*) are located around the first ATG (A, adenine; T, thymine; G, guanine; C, cytosine) of region S. It should be noted that the heptanucleotide sequence, 5′ GCATGGG, is repeated at both extremities of this additional sequence and could be related to its insertion or deletion.

The gene C (nucleotides 1901 to 2450) codes for P19, the major polypeptide of the viral core, as shown by expression of HBcAg in *Escherichia coli* (*16, 37*). Neither genes S or C appear to contain any introns (*15–17*).

The region P (nucleotides 2307 to 1620) is particularly long, since it covers more than 75 percent of the genome. It may code for a basic polypeptide of about 90,000 daltons, which is unusually rich in histidine. As DNA polymerases are of comparable molecular weight, the hypothesis that this region could be the gene for the virion DNA polymerase has been raised (*15–17*). The comparison of the sequences of the region P in Table 1 warrant a comment. Whereas the divergence in amino acids is 16.6 percent in the amino terminal part, it is 4.9 percent in the central part and 3.0 percent in the carboxyl terminal part, where there is no constraint due to the overlap of two coding regions. The relatively good conservation of the carboxyl terminal part could correspond to functionally important domain.

Table 1. Comparison of the translation of the L strand of two HBV genomes (*67*). The amino acid sequences deduced from the L strand in three reading frames were compared for the two complete nucleotide sequences of Galibert *et al.* (*15*) and Valenzuela *et al.* (*17*) after excluding the additional sequences. The amino acid differences were scored for significant changes, such as nonpolar to basic. The L strand was divided into four segments consisting of the known genes S and C, and two intervening segments. The regions pre-S, S, C, P, and X, as defined in Fig. 3, are shown in each of the three reading frames for the respective segments.

Nucleotide numbers in segment	Reading frame region	Amino acid difference (%)
2451 to 154	P	16.6
	Pre-S	11.7
	Phase 3	22.0
155 to 833	P	4.9
	S	2.2
	Phase 3	6.6
834 to 1900	P	3.0
	Phase 2	10.7
	X	3.9
1901 to 2450	Phase 1	13.1
	C	1.1
	Phase 3	13.1

The region X (nucleotides 1374 to 1835) could code for a basic polypeptide of 154 amino acids. The nick of the L strand seems to be located in this region that cannot be transcribed completely from the DNA present in the virion. This does not eliminate a possible coding function since the molecular structure of the DNA supporting the viral transcription is not defined.

Different sequences which may correspond to signals regulating transcription and translation were detected on the viral DNA (*15–17*). These sequences include the TATA box observed in front of the transcribed region (*41*), the AAUAAA-like (U, uracil) sequence found at the 3′ end of mRNA near the poly(A) tail (*42*), and the sequence complementary to the 3′ end of 18*S* rRNA (ribosome RNA) (*43*). Valenzuela *et al.* have reported TATA boxes and AAUAAA-like sequences in relation to the different open regions on the L strand (*17*). The precise AAUAAA sequence is not found. Pasek *et al.* have reported a stem with a loop that could pair with the 3′ end of 18*S* rRNA, located immediately before gene C (*16*). In the absence of experimental data, the interpretation of these sequences is partly speculative. Analysis of viral mRNA's is necessary for understanding more precisely the genetic organization.

The four coding regions described cover nearly the entire genome (except for 63 nucleotides). The region P overlaps the three other regions and entirely covers the region S. The ability of a DNA sequence to code two different polypeptides in two overlapping reading

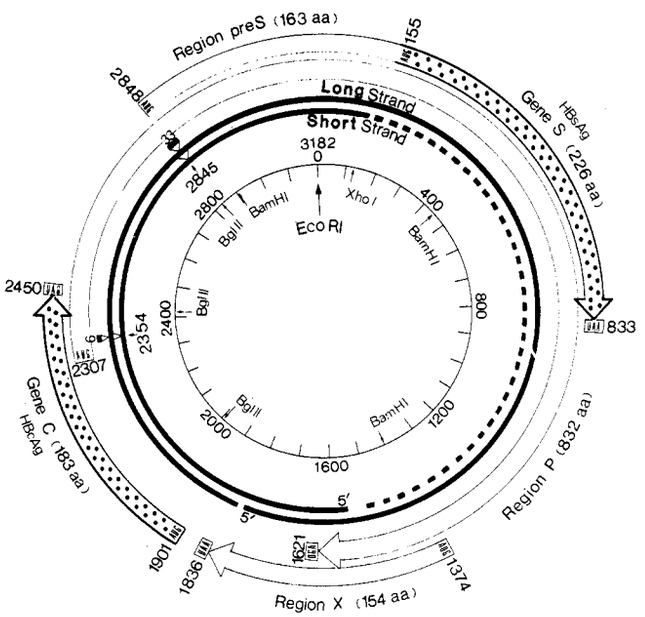

Fig. 3. Physical structure and proposed genetic organization of the HBV genome. The 5′ end of the L strand is base-paired with the 5′ end of the S strand. The dashed line corresponds to the variable single-stranded region. Certain restriction sites indicated by arrows correspond to the physical map of the HBV/ay genome (*35*). The single Eco RI site is used as the point of origin in the physical map, and the single Xho I site is indicated in the conventionally clockwise orientation of numbering the nucleotides from 0 to 3182. The two sequences of 6 and 33 nucleotides at positions 2354 and 2845 are those present in the HBV/ad clone analyzed by Valenzuela *et al.* (*17*). The broad arrows surrounding the genome correspond to the four large open regions of the L strand transcript, conserved in the three recorded sequences (*15–17*). These four potential coding regions are called region S (divided into pre-S and gene S), P, X, and C. The number of amino acids in the brackets corresponds to the length of the hypothetical polypeptide. The two regions corresponding to the defined genes S and C are represented by stippling.

frames has been reported for other viruses (*44*). However, in the case of HBV this capacity is extensively used since the L strand could be read one and a half times. It is likely that the extremely small size of the genome (the smallest of the known animal DNA viruses) dictates this type of organization to make available a larger coding capacity. The constraints of size and of conservation of the reading frames may explain the observation of only two insertions or deletions encountered in the three sequences. This is in contrast to the large number of point mutations. In both cases, the number of additional nucleotides is a multiple of three, and hence the reading frame is unaltered.

Biology of HBV and Experimental Models

Natural or experimental transmission of HBV has been achieved only in humans and chimpanzees. Attempts to propagate HBV in tissue cultures and numerous species of experimental animals have been unsuccessful (*45*). Such a narrow host specificity and lack of an in vitro system for propagation of HBV has impeded understanding the biology of HBV. Although transmission studies in humans and chimpanzees have resulted in serological and histological characterization of a spectrum of clinical manifestations of HBV infection, the biology of the virus itself and the mechanisms involved in assembly of the elements of the virion have remained enigmatic.

Expression of the viral genome has been studied in different in vitro systems. The first consists of a human hepatoma cell line (PLC/PRF/5 line) established by Alexander and colleagues (*46*). The cells synthesize and excrete in the medium only 22-nanometer particles with all of the immunochemical characteristics of HBsAg (*47*). The Southern blot technique (*48*) has been used to demonstrate that the whole HBV DNA is present in the cellular DNA only in an integrated form (*49–52*). In the total RNA of these cells, Marion *et al.* (*49*) have demonstrated sequences hybridizing with the whole genome (*49*). Using the blotting technique of Alwine *et al.* for RNA hybridization (*53*), Chakraborty *et al.* (*50*) have detected poly(A)-containing RNA of 2500 and 3050 nucleotides whereas Edman *et al.* (*52*) have found only one transcript of 2000 nucleotides, which hybridizes to the region S of the genome. In view of these conflicting results, further analyses of the viral RNA are needed.

Another approach to study expression of the viral genes consists of introducing the previously cloned genome, or its fragments, into cell lines. Dubois *et al.* (*54*) reported the introduction of HBV DNA into the mouse L cell line using the method of DNA-mediated gene transfer (*55*). This procedure resulted in (i) integration of HBV DNA into the cellular DNA and (ii) secretion of HBsAg particles into the medium without lysis of the cells. Other viral antigens were not detected. The reason why only the gene S is expressed by the L cell line despite the presence of an entire genome is still unclear. Moriarty *et al.* constructed a hybrid of SV40 DNA and a fragment of HBV DNA which when introduced into a monkey kidney cell line induced synthesis of HBsAg particles (*56*). Hirschman and colleagues have transfected HeLa cells with a cloned and recircularized HBV genome which resulted in synthesis of both HBsAg and HBcAg (*57*). Dane-like particles in addition to HBsAg particles were observed in the cell-free supernatants suggesting that the cells produce an entire virion.

The etiologic role of HBV in manifestation of hepatoma has been indicated by numerous epidemiologic observations (*58*). The capacity of HBV to integrate its DNA into the host genome had long been suspected. Indeed, cell transformation by the known oncogenic viruses is generally accompanied by integration of at least a part of their genome into cellular DNA (*59*). Integration of HBV DNA in the cellular DNA of hepatoma tissue and cell line PLC/PRF/5 has been demonstrated (*49–52*) by the Southern blotting method. However, this is not sufficient evidence for the direct oncogenic role of the virus. Integration of viral DNA was similarly observed in nontumorous portions of the liver of hepatoma patients (*60*). Shouval *et al.* have shown that in two chimpanzees with chronic hepatitis HBV sequences do not appear to be integrated (*61*). Nevertheless, Brechot and colleagues have studied 17 cases of chronic hepatitis type B in man by the Southern blotting method and have revealed the existence of integrated sequences (*60*).

The foregoing observations raise two issues.

1) In order to decide whether there is a difference in the way the viral DNA is integrated in the case of either chronic hepatitis or hepatoma, it is necessary to make a precise comparison of sequences integrated into DNA of the affected and unaffected parts of the liver in the same patients. Cloning of the integrated viral sequences in different clinical conditions should yield the requisite data. Such data would also confirm the suggestion by Edman *et al.* (*52*) that the viral genome is integrated at the nick and would also possibly elucidate the mechanism of its insertion.

2) As to integration of viral DNA into the host DNA essential for the viral cycle, as hypothesized by Hirschman (*62*), we are likely to observe integrated viral sequences in all of the clinical situations, including acute hepatitis. However, if that is not the case, the integration could constitute a way of persistence of the virus and a characteristic of chronic infection. The results obtained with the cell line PLC/PRF/5 and with the mouse L cells transformed by HBV DNA indicate that at least one of the viral genes, the gene S, may be expressed at a high level with only one or a few copies of the gene integrated into the cellular DNA. By analogy we can imagine that the persistence of HBsAg production and possibly other viral antigens in hepatocytes of certain patients with chronic hepatitis could be due to the presence of integrated viral sequences and would not necessitate the presence of free viral genomes or viral replication. In such patients antiviral therapy is likely to be ineffective.

Conclusions

Our current knowledge of the biology of HBV has developed from the investigations in seroepidemiology, immunochemistry of viral antigens (especially HBsAg), and the molecular biology of the virus. Development of serological assays useful in clinical diagnosis, public health, and epidemiology has been fulfilled. Although the immunopathological mechanisms underlying the feature of large amounts of HBsAg in the blood of chronic carriers are not yet clarified, the isolation of the viral envelope from plasma has led to preparation of a safe and effective vaccine (*8*). Peptide analogs of the antigenic determinants of HBsAg could lead to the future development of synthetic vaccines (*18*).

The limitation imposed by unavailability of an in vitro system for propagation of the virus has been partially overcome by the cloning of the genome in *E. coli* and the success in transfection of mammalian cell lines with cloned HBV DNA. From these results we can anticipate rapid progress in our understanding of the mechanisms of viral gene expression. Occurrence of HBV-like viruses in woodchucks and squirrels (*63*) and in ducks (*64*) has opened up the prospects

for suitable laboratory models potentially useful in studying the host virus interactions. The existence of a large class of evolutionarily related DNA viruses including HBV, HBV-like nonhuman viruses, and eventually human hepatitis viruses that are neither A nor B is more than likely. This class of viruses has been termed hepadna (*65*).

The use of cloned HBV DNA in a hybridization technique could provide a sensitive new diagnostic tool for the detection of HBV DNA in the liver or in the serum. Indeed, some data demonstrate the integration of HBV sequences in the hepatocellular DNA both in hepatoma and in chronic hepatitis (*60*). These integrated sequences can further elucidate the molecular mechanisms of integration and the possible role of HBV in cell transformation. The synthesis of HBsAg from cloned HBV DNA in both eukaryotic and prokaryotic (*66*) systems could possibly provide alternative forms of HBsAg, which may be more economical and safer than the current vaccines.

References and Notes

1. W. Szmuness, *Am. J. Pathol.* **81**, 629 (1975).
2. W. S. Robinson, *Annu. Rev. Microbiol.* **31**, 357 (1977).
3. G. N. Vyas, A. B. Ibrahim, K. R. Rao, R. Schmid, *Life Sci.* **15**, 261 (1974); C. E. Stevens, R. P. Beasley, J. Tsui, W. E. Lee, *N. Engl. J. Med.* **292**, 771 (1975); D. Y. Chien and G. N. Vyas, *ibid.* **299**, 1253 (1978); C. E. Stevens, R. A. Neurath, R. P. Beasley, W. Szmuness, *J. Med. Virol.* **3**, 237 (1979); S. Krugman, L. R. Overby, I. K. Mushahwar, C.-M. Ling, G. G. Froesner, F. Dienhardt, *N. Engl. J. Med.* **300**, 101 (1979).
4. G. N. Vyas, S. N. Cohen, R. Schmid, Eds., *Viral Hepatitis* (Franklin Institute Press, Philadelphia, 1978).
5. B. S. Blumberg, H. J. Alter, S. Visnich, *J. Am. Med. Assoc.* **191**, 541 (1965).
6. D. Dane, C. H. Cameron, M. Briggs, *Lancet* **1977-I**, 357 (1977).
7. J. Drouet *et al.*, *Biomedicine* **22**, 158 (1975).
8. R. H. Purcell and J. L. Gerin, *Am. J. Med. Sci.* **270**, 395 (1975); M. R. Hilleman, E. B. Buynak, R. R. Roehm, A. A. Tytell, A. U. Bertland, G. P. Lampson, *ibid.*, p. 401; P. Maupas, P. Coursaget, A. Goudeau, J. Drucker, *Lancet* **1976-I**, 1367 (1976); A. J. Zuckerman, *Nature (London)* **272**, 579 (1978); W. Szmuness *et al.*, *N. Engl. J. Med.* **303**, 833 (1980).
9. G. L. Le Bouvier and A. Williams, *Am. J. Med. Sci.* **270**, 165 (1975).
10. G. N. Vyas, M. A. Mason, E. W. Williams, in *Hepatitis and Blood Transfusion*, G. N. Vyas, H. A. Perkins, R. Schmid, Eds. (Grune & Stratton, New York, 1972), p. 144; A. M. Courouce and J. P. Soulier, *Bibl. Haematol. (Basel)* **42**, 31 (1976); A. M. Courouce and P. V. Holland, in (*4*), p. 649. Determinant a was originally subdivided into a_1, a_2, and a_3. Because a did not have subspecificities in association with r, the subspecificities were assigned to w instead of the a determinant. This is incongruent with demonstrated serologic heterogeneity of the a determinant.
11. G. N. Vyas, E. W. Williams, G. G. B. Klaus, H. E. Bond, *J. Immunol.* **108**, 1114 (1972); Y. Sanchez, I. Ionescu-Matiu, F. B. Hollinger, J. L. Melnick, G. R. Dreesman, *J. Gen. Virol.* **48**, 273 (1980); J. Koistenen, *J. Virol.* **35**, 20 (1980); H. Shirashi, K. Akiyama, N. Ishida, *J. Gen. Virol.* **48**, 31 (1980).
12. D. L. Peterson, D. Y. Chien, G. N. Vyas, D. Nitecki, H. E. Bond, in (*4*), p. 569.
13. P. Valenzuela, P. Gray, M. Quiroga, J. Zaldivar, H. M. Goodman, W. J. Rutter, *Nature (London)* **280**, 815 (1979).
14. P. Charnay, E. Mandart, A. Hampe, F. Fitoussi, P. Tiollais, F. Galibert, *Nucleic Acid Res.* **7**, 335 (1979).
15. F. Galibert, E. Mandart, F. Fitoussi, P. Tiollais, P. Charnay, *Nature (London)* **281**, 646 (1979).
16. M. Pasek *et al.*, *ibid.* **282**, 575 (1979).
17. P. Valenzuela, M. Quiroga, J. Zaldivar, P. Gray, W. J. Rutter, in *Animal Virus Genetics*, B. Fields, R. Jaenisch, C. F. Fox, Eds. (Academic Press, New York, 1980), p. 57.
18. G. N. Vyas, D. R. Milich, T. Gottfried, D. Nitecki, and M. Karels reported in July 1980 the feasibility of a synthetic vaccine (IV International Immunology Congress, Paris, Abstr.). A computer program based on the Fasman scheme was used by M. Karels (Department of Molecular Biology, University of California, Berkeley). The most hydrophilic region (110 to 160) was expected to contain the major antigenic determinant of lysine-dependent activity of HBsAg [K. R. Rao and G. N. Vyas, *Microbios* **9**, 239 (1974)]. Synthesis of oligopeptide analogs in this region (for example, residues 134 to 146) yielded antigenically active peptides with the common a determinants, potentially useful in new vaccines against HBV infection [G. N. Vyas, in *Hepatitis B Vaccine*, P. Maupas and P. Guesry, Eds. (Elsevier, Amsterdam, 1981)]. Synthesis of peptides eliciting antibodies reacting with the native envelope protein of Dane particles has been recently reported by R. A. Lerner, N. Green, H. Alexander, F. T. Liu, J. G. Sutcliff, and T. M. Shinnick (*Proc. Natl. Acad. Sci. U.S.A.*, in press).
19. H. Shiraishi, T. Kohama, R. Shirachi, N. Ishida, *J. Gen. Virol.* **36**, 207 (1977).
20. C. J. Burrell, E. Proudfoot, G. A. Keen, B. P. Marmion, *Nature (London)* **243**, 260 (1973); R. Chairez, S. Steiner, J. L. Melnick, G. R. Dreesman, *Intervirology* **1**, 224 (1973); J. W. Shih and J. L. Gerin, *J. Immunol.* **115**, 634 (1975); C. J. Burrell, G. Leadbetter, P. Mackay, B. P. Marmion, *J. Gen. Virol.* **33**, 41 (1976); D. L. Peterson, I. M. Roberts, G. N. Vyas, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1530 (1977); H. Shiraishi, R. Shirachi, N. Ishida, T. Sekine, *J. Gen. Virol.* **38**, 363 (1978); D. L. Peterson, *J. Biol. Chem.*, in press.
21. M. Z. Atassi, *Arch. Biochem. Biophys.* **120**, 56 (1967).
22. J. Skelly, C. R. Howard, A. J. Zuckerman, *J. Gen. Virol.* **41**, 447 (1978).
23. S. Mishiro *et al.*, *J. Immunol.* **124**, 1589 (1980).
24. G. N. Vyas, K. R. Rao, A. B. Ibrahim, *Science* **178**, 1300 (1972); N. Sukeno, R. Shirachi, J. Yamaguchi, N. Ishida, *J. Virol.* **9**, 182 (1972); G. R. Dreesman, F. B. Hollinger, R. M. MacCombs, J. L. Melnick, *J. Gen. Virol.* **19**, 129 (1973).
25. For review, see J. L. Gerin and J. W. Shih, in (*4*), p. 147.
26. M. Nakano, T. Tohita, T. Ando, *Int. J. Pep. Protein Res.* **8**, 565 (1976).
27. C. Albin and W. S. Robinson, *J. Virol.* **34**, 297 (1980).
28. K. C. Lam, M. J. Tong, J. Rakela, *Infect. Immun.* **16**, 403 (1977); K. Takahashi *et al.*, *J. Immunol.* **122**, 275 (1979); A. R. Neurath and N. Strick, *J. Gen. Virol.* **42**, 646 (1979).
29. A. R. Neurath and M. Strick, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1702 (1977); G. N. Vyas, D. L. Peterson, R. M. Townsend, S. R. Damle, L. O. Magnius, *Science* **198**, 1068 (1977); K. Takahashi, M. Imai, Y. Miyakawa, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1952 (1978).
30. C. Trepo, O. Hantz, L. Vitvitski, P. Chevalier, A. Williams, J. M. Lemaire, in (*4*), p. 203. M. Mayumi and Y. Miyakawa reported at the 1981 International Symposium on Viral Hepatitis, New York, 2 April 1981, that P19 carries HBcAg and two distinct determinants of HBeAg termed e_a and e_b.
31. M. M. Rizetto *et al.*, *Gut* **18**, 997 (1977); M. Rizetto, J. W. Shih, D. J. Cocke, R. H. Purcell, G. Verme, J. L. Gerin, *Lancet* **1979-II**, 896 (1979); M. Rizetto, M. G. Canese, J. L. Gerin, W. T. London, D. L. Sly, R. H. Purcell, *J. Infect. Dis.* **141**, 590 (1980); M. Rizetto, J. W. Shih, J. L. Gerin, *J. Immunol.* **125**, 318 (1980); M. Rizetto, B. Hoyer, M. G. Canese, J. Wai-Kuo-Shih, R. H. Purcell, J. L. Gerin, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6124 (1980).
32. M. Rizetto, M. H. Purcell, J. L. Gerin, *Perspect. Virol.*, in press.
33. P. Kaplan, R. L. Greenman, J. L. Gerin, R. H. Purcell, W. S. Robinson, *J. Virol.* **12**, 995 (1973); W. S. Robinson, D. A. Clayton, R. L. Greenman, *ibid.* **14**, 384 (1974); J. F. Hruska, D. A. Clayton, J. L. R. Rubenstein, W. S. Robinson, *ibid.* **21**, 666 (1977).
34. J. Summers, A. O'Connell, I. Millman, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4597 (1975).
35. A. Fritsch, C. Pourcel, P. Charnay, P. Tiollais, *C. R. Acad. Sci. Paris Ser. D* **287**, 1453 (1978); P. Charnay, C. Pourcel, A. Fritsch, A. Louise, P. Tiollais, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2222 (1979).
36. A. Siddiqui, F. Sattler, W. S. Robinson, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4664 (1979).
37. C. J. Burrell, P. Mackay, P. J. Greenway, P. H. Hofschneider, K. Murray, *Nature (London)* **279**, 43 (1979).
38. J. J. Sninsky, A. Siddiqui, W. S. Robinson, S. N. Cohen, *ibid.*, p. 346.
39. F. Sattler and W. S. Robinson, *J. Virol.* **32**, 226 (1979).
40. W. H. Gerlich and W. S. Robinson, *Cell* **21**, 801 (1980).
41. E. B. Ziff and R. M. Evans, *ibid.* **15**, 1463 (1978); F. Bannon *et al.*, *Nature (London)* **278**, 428 (1979).
42. H. J. Proudfoot and G. G. Brownlee, *Nature (London)* **263**, 211 (1976); A. Efstratiadis and F. C. Kafatos, *Cell* **10**, 571 (1977); N. J. Proudfoot, *ibid.*, p. 559.
43. F. E. Baralle and G. G. Brownlee, *Nature (London)* **274**, 84 (1978); O. Hagenbuchle, M. Santer, J. A. Steitz, *Cell* **13**, 551 (1978); P. Lomedico, N. Rosenthal, A. Efstratiadis, W. Gilbert, R. Kolodner, R. Tizard, *ibid.* **18**, 545 (1979).
44. F. Sanger *et al.*, *Nature (London)* **265**, 687 (1977); B. G. Barrell, D. C. Shaw, J. E. Walker, F. D. Northrop, G. N. Godson, J. C. Fiddes, *Biochem. Soc. Trans.* **6**, 63 (1978); B. Reddy *et al.*, *Science* **200**, 494 (1978); W. Fiers *et al.*, *Nature (London)* **273**, 113 (1978).
45. S. Krugman, in (*4*), p. 3.
46. J. J. Alexander, E. M. Bay, E. W. Geddes, G. Lecatsas, *S. Afr. Med. J.* **50**, 1124 (1976).
47. P. L. Marion, F. H. Salazar, J. J. Alexander, W. S. Robinson, *J. Virol.* **32**, 796 (1979).
48. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
49. P. L. Marion, F. H. Salazar, J. J. Alexander, W. S. Robinson, *J. Virol.* **33**, 795 (1980).
50. P. R. Chakraborty, N. Rinz-Opazo, D. Shouval, D. A. Shafritz, *Nature (London)* **286**, 531 (1980).
51. C. Brechot, C. Pourcel, A. Louise, B. Rain, P. Tiollais, *ibid.*, p. 533.
52. J. Edman, P. Gray, P. Valenzuela, L. B. Rall, W. J. Rutter, *ibid.*, p. 535.
53. J. C. Alwine, D. J. Kemp, G. R. Stark, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5350 (1977).
54. M. F. Dubois, C. Pourcel, S. Rousset, C. Chany, P. Tiollais, *ibid.* **77**, 4549 (1980).
55. M. Wigler *et al.*, *Cell* **16**, 777 (1979).
56. A. M. Moriarty, B. H. Hoyer, J. W. Shih, J. L. Gerin, D. H. Hamer, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2606 (1981).
57. S. Z. Hirschman, P. Price, E. Garfinkel, J. Kristman, G. Acs, *ibid.* **77**, 5507 (1980).
58. W. Szmuness, *Prog. Med. Virol.* **24**, 40 (1978).
59. For review, see R. A. Weinberg, *Annu. Rev. Biochem.* **49**, 197 (1980).
60. C. Brechot *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
61. D. Shouval *et al.*, *ibid.* **77**, 6147 (1980).
62. S. Z. Hirschman, *Lancet* **1975-II**, 436 (1975).
63. J. Summers, J. M. Smolec, R. Snyder, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4533 (1978); B. G. Werner, J. M. Smolec, R. Snyder, J. Summers, *J. Virol.* **32**, 314 (1979); I. W. Cummings *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1842 (1980); P. L. Marion, L. S. Oshiro, D. C. Regnery, G. H. Scullard, W. S. Robinson, *ibid.*, p. 2941.
64. W. S. Mason, G. Seal, J. Summers, *J. Virol.* **36**, 829 (1980).
65. W. S. Robinson, *Ann. N.Y. Acad. Sci.* **354**, 371 (1980).
66. P. Charnay, M. Gervais, A. Louise, F. Galibert, P. Tiollais, *Nature (London)* **286**, 893 (1980).
67. The data analyses on the nucleotide sequences and amino acid sequences were performed at the computer center of the Institut Pasteur.

68. We thank Dr. R. Acher for critical discussions, Dr. C. Brechot for unpublished results, Dr. W. Rutter for preprint of the complete nucleotide sequence of Valenzuela *et al.* (*17*), Dr. P. Kourilsky and Dr. M. Yaniv for an intramural review of the manuscript, and Dr. C. Pourcel and Dr. S. Wain-Hobson for critical commentary, and the Fullbright Foundation for partial support of Dr. Vyas.