The EMBO Journal vol.4 no.6 pp.1507–1514, 1985

Transcription of woodchuck hepatitis virus in the chronically infected liver

Tarik Möröy, Jeanne Etiemble, Christian Trépo[1], Pierre Tiollais and Marie-Annick Buendia

Unité de Recombinaison et Expression Génétique (INSERM U163, CNRS LA 271), Institut Pasteur, 28 rue du Dr Roux, 75015 Paris, and [1]Unité de Recherches sur les Hépatites, (INSERM U.271), UER Alexis Carel, rue Guillaume Paradin, 69008 Lyon, France

Communicated by P.Tiollais

The transcription of woodchuck hepatitis virus (WHV) genome was studied in the liver of chronically infected woodchucks by Northern blot, nuclease mapping and primer extension analysis. Two major transcripts, 2.1 and 3.7 kb in length, and several minor transcripts were found in samples which supported active WHV replication. The 2.1-kb RNA represents the major transcript of the S gene, encoding the viral surface antigen (WHsAg) as demonstrated by blot-hybridization experiments. Two transcription initiation sites were localized downstream of the second AUG of the pre-S region, 139 and 152 nucleotides upstream of the translation initiation codon of the S gene. The 3.7-kb transcript, present in an equal amount, is slightly larger than the WHV genome and could be involved in the expression of all viral proteins. The data derived from RNA mapping strongly suggest that this transcript is initiated ~70 nucleotides upstream of the C gene, encoding the viral core antigen (WHcAg), and represents the message for WHcAg. It might also serve in the viral replication cycle as a potential template for reverse transcription. All WHV-specific transcripts were found to be processed at a unique site, 20 nucleotides downstream of the polyadenylation signal situated within the core gene. A different set of WHV-specific mRNAs was observed in a woodchuck hepatocellular carcinoma when only integrated forms of WHV DNA could be detected. Two RNA species of 2.3 and 4.6 kb were characterized. The 3.7-kb RNA was absent, reinforcing the hypothesis that this transcript corresponds to the pre-genome.

Key words: WHV transcription/infected liver/mRNAs/nuclease mapping

Introduction

Woodchuck hepatitis virus (WHV) was originally described by Summers *et al.* (1978) in association with liver diseases in the Eastern woodchuck (*Marmota monax*). WHV belongs to the group of animal DNA viruses recently designated as 'hepadna viridae' (Robinson *et al.*, 1982), including also human hepatitis B virus (HBV), ground squirrel hepatitis virus (GSHV) and duck hepatitis B virus (DHBV). They share a common virion structure, genetic organization of viral DNA and the ability to cause persistent infections (Marion *et al.*, 1980; Mason *et al.*, 1980; Summers, 1981; Feitelson *et al.*, 1981; Galibert *et al.*, 1982; Mandart *et al.*, 1984; Seeger *et al.*, 1984). The genome of the 'hepadna viridae' is a small circular and partly double-stranded DNA molecule. A long or L(−) strand of fixed length (3.3 kb for WHV) with a nick or a gap of a few nucleotides and a short or S(+) strand of variable length maintain the circular structure by base pairing of their 5′ ends (for a review, see Tiollais *et al.*, 1984). Four large open reading frames termed S, C, P and X, which are conserved among the mammalian 'hepadna' viruses, are located on the L(−) strand. The S region is subdivided into the pre-S region and the S gene which corresponds to the coding sequence of the surface antigen. The C region encodes the core antigen, a structural polypeptide of the virion. A protein encoded by the region X still has to be defined. The region P probably corresponds to the gene for the endogenous DNA polymerase associated with the virion. The discovery of a reverse transcriptase activity of the viral DNA polymerase allowed the development of a model for the viral replication cycle that involves a RNA intermediate called pre-genome (Summers and Mason, 1982).

The pathologic effects of WHV infection in woodchucks, i.e., acute hepatitis, chronic persistent and chronic active hepatitis, are very similar to HBV-induced diseases in man (Frommel *et al.*, 1984). In addition, as for HBV, chronic active hepatitis is frequently associated with the development of hepatocellular carcinoma (Snyder and Summers, 1980; Gerin, 1984). This pathology has never been reported for ground squirrels, Pekin ducks or HBV-infected chimpanzees. Therefore, infection of woodchucks by WHV is likely to represent the best animal system for studying the relationship between HBV infection and the appearance of hepatocellular carcinoma (HCC).

Precise studies on the transcription of HBV have been hampered until now by the lack of a cell culture system as well as by the paucity of liver material from humans. Several models have been used, including permanent cell lines established from human hepatocellular carcinoma (Chakraborty *et al.*, 1980; Edman *et al.*, 1980; Pourcel *et al.*, 1982), animal cells transfected with cloned HBV DNA (Gough and Murray, 1982; Pourcel *et al.*, 1982; Gough, 1983; Cattaneo *et al.*, 1983) or with recombinants between heterologous viral promoter-sequences and subgenomic fragments of HBV DNA (Simonsen and Levinson, 1983; Laub *et al.*, 1983; Standring *et al.*, 1984; Will *et al.*, 1984; Michel *et al.*, 1984). HBV transcripts have also been identified in the liver of an infected chimpanzee (Cattaneo *et al.*, 1983, 1984). In all these models the S gene is efficiently expressed and the HBsAg transcriptional unit has been mapped on the HBV genome, including the localization of two different promoters and a unique polyadenylation signal, but the question of a splicing event occurring in the pre-S region is still unanswered (Galibert *et al.*, 1982; Gough, 1983; Laub *et al.*, 1983).

The transcription of the other viral genes is so far less documented. Results concerning the core gene have been obtained by '*in vitro*' transcription assays of cloned HBV DNA (Rall *et al.*, 1983) or nuclease mapping and blot hybridization analysis of HBV-specific RNA in different systems (Gough, 1983; Standring *et al.*, 1984; Will *et al.*, 1984; Cattaneo *et al.*, 1984). However, these models, with the exception of the infected chimpanzee, do not allow viral replication and therefore do not mimic

© IRL Press Limited, Oxford, England.

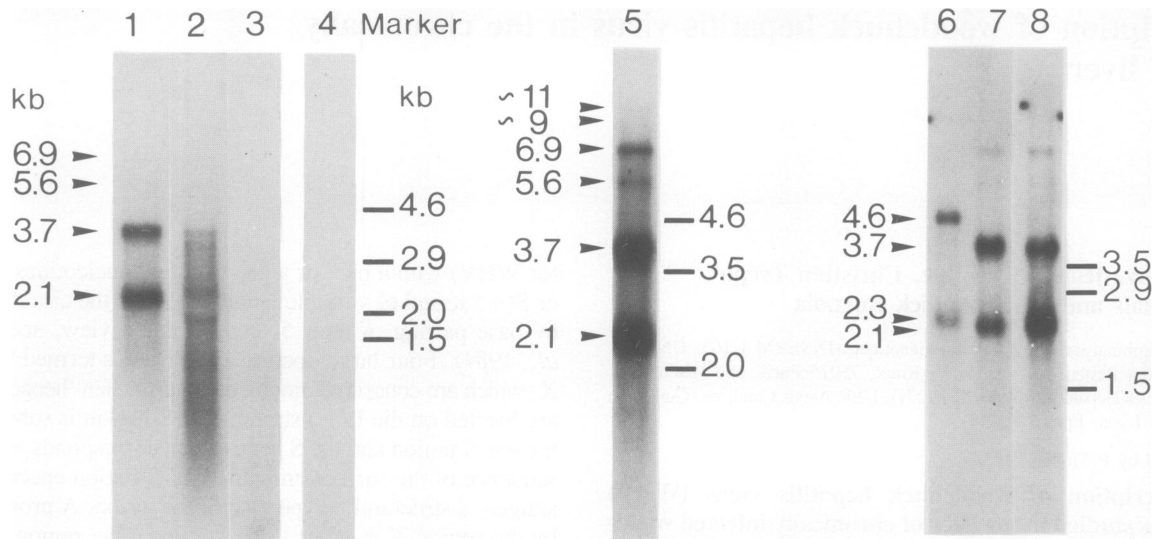

Fig. 1. Analysis of WHV-related transcripts in the liver of chronically infected woodchucks. Poly(A)$^+$ and poly(A)$^-$ RNA were prepared and selected as described in Materials and methods. About 7 μg of poly(A)$^+$ RNA and 50 μg of poly(A)$^-$ RNA were routinely separated on formaldehyde-agarose gels and subsequently transferred to nitrocellulose filters. WHV-related transcripts were identified by hybridization with the radiolabelled 3.3-kb WHV-specific *Eco*RI fragment of pBH20-WHV-1. The RNA size markers were mammalian 28S and 18S rRNA, *E. coli* 23S and 16S rRNA and bacteriophage MS2 RNA, visualized by ethidium bromide staining. The liver samples tested were: a tumor-free chronic carrier, woodchuck W64 (February 1984) (**lane 1**: poly(A)$^+$ RNA; **lane 2**: poly(A)$^-$ RNA; **lane 4**: same sample as in **lane 1**, probed with pBR322 DNA); the uninfected woodchuck W101 (**lane 3**: 50 μg of total RNA); the chronically infected W78 woodchuck with hepatocellular carcinoma [non-tumorous liver, poly(A)$^+$ RNA (**lane 5**)]; two other tumor-bearing animals: poly(A)$^+$ RNA from W64 tumorous liver (July 1984) (**lane 6**); W64 non-tumorous tissue (**lane 7**); and W74 tumorous tissue (**lane 8**). The blots were exposed for 8–24 h with an intensifying screen at −70°C.

a virus-infected liver.

In this study, we characterized WHV transcripts by blot hybridization as well as by nuclease mapping and primer extension experiments. In addition, WHV-specific transcripts were identified in two different cases of hepatocellular carcinoma.

Results

Transcription of WHV in infected woodchuck livers

WHV transcripts were investigated in biopsy samples of chronically infected woodchucks, either from the liver of tumor-free chronic carriers, or from tumorous and non-tumorous regions of the liver of HCC-bearing animals. A general overview of the size of WHV transcripts was obtained by gel electrophoresis and hybridization analysis of polyadenylated and non-polyadenylated RNA. Figure 1 shows that the poly(A)$^+$ RNA from tumor-free liver and from the non-tumorous part of carcinoma-carrying liver contain WHV-specific transcripts identical in number and size. Two predominant poly(A)$^+$ RNA had an estimated length of 3.7 and 2.1 kb, by reference to RNA size markers on several different gels. They represented 40 and 45%, respectively, of total WHV transcripts when the autoradiograms were scanned on a Beckman scanner (data not shown). Two minor poly(A)$^+$ species of 6.9 and 5.6 kb were also detected (Figure 1), representing ~10% of the WHV transcripts. Other minor bands of 11 and 9 kb were only occasionally observed. A smear of RNA smaller than 3.7 kb appeared in the poly(A)$^-$ tracks, sometimes accompanied by one or two discrete bands in the 1.0–2.0 kb area (Figure 1, lane 2). The analysis of total RNA of non-infected woodchuck liver as well as the use of ^{32}P-labelled pBR322 DNA to probe the blots shown in Figure 1 did not allow the detection of any signal (Figure 1, lanes 3,4). The total WHV-specific transcripts were estimated to represent ~0.2% of total poly(A)$^+$ RNA in the liver of chronically infected woodchucks by dot-blot experiments (data not shown).

Investigation of WHV-specific RNA from two independent hepatocellular carcinoma samples showed two different situations. In the tumor from W74 woodchuck, the pattern of poly(A)$^+$ RNA was indistinguishable from those previously shown for tumor-free liver (Figure 1, lanes 7,8). In a second tumor (W64) the WHV-specific transcripts appeared to be entirely different in size. Two species of 4.6 kb and 2.3 kb were present in nearly equal amounts, whereas the above-described WHV transcripts were not detectable, even after long exposure of the blots.

To determine the transcribed regions of the WHV genome, five probes corresponding to definite fragments of the viral DNA were prepared as described in Materials and methods. The location of these probes is illustrated in the lower part of Figure 2. Hybridization patterns of poly(A)$^+$ RNA from chronically infected liver showed that the RNA species of 3.7 kb gave a positive signal with all probes, covering the entire WHV genome (Figure 2, upper part, lanes N). The 2.1-kb transcript hybridized only with probes I and II, which cover the WHV genome from nucleotides 1 to 2190 on the physical map (Figure 2, lower part), and not with probes III, IV, V or with a *Bgl*II-*Eco*RI probe (773 nucleotides in length) prepared from pHB20-WHV-1 (data not shown). Thus, it is likely to correspond to the WHsAg mRNA. The analysis of poly(A)$^+$ RNA prepared from woodchuck W64 tumorous tissue was carried out in the same experiment (Figure 2, upper part, lanes T). The 4.6-kb transcript hybridized with all probes, with the exception of probe III, which covers the C gene region (nucleotides 2190–2535). The hybridization patterns of the 2.3-kb transcript were identical to those of the 2.1-kb RNA present in non-tumorous liver.

Southern blot analysis of viral DNA

The discrepancies observed between the WHV poly(A)$^+$ RNA patterns of woodchuck W64 hepatocarcinoma and the other patterns obtained from tumorous or non-tumorous liver tissues might be explained by a different state of WHV DNA in the samples. The presence of WHV DNA in W64 tumorous and non-tumorous tissues and in W74 tumor samples was investigated using the tech-

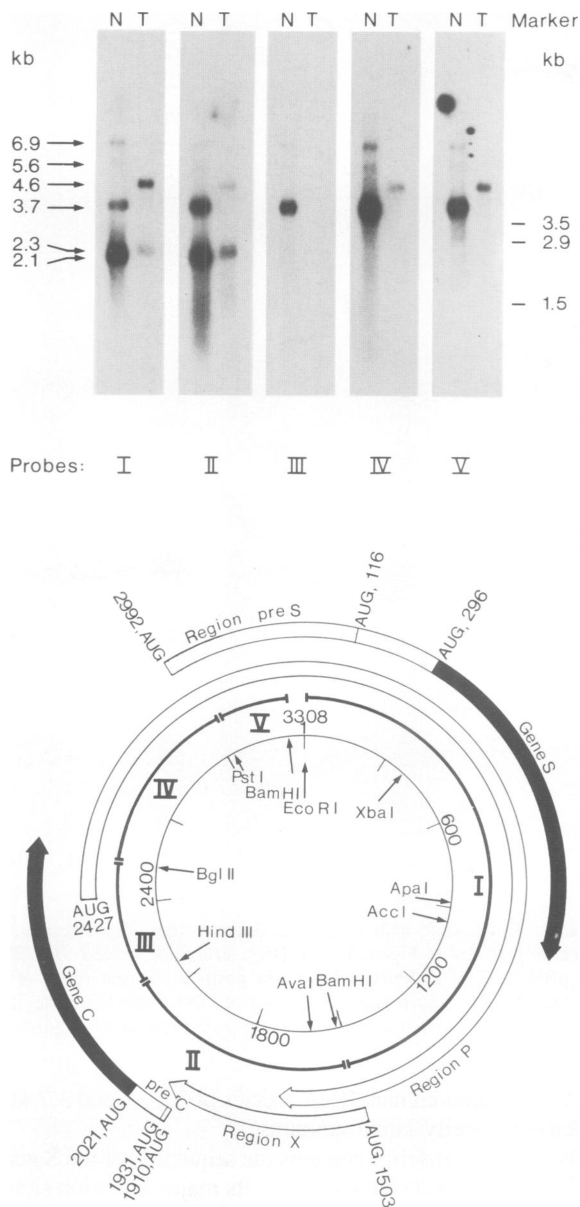

Fig. 2. Hybridization of WHV-specific poly(A)$^+$ RNA derived from non-tumorous tissue of W64 **(lanes N)** or from tumorous tissue of W64 **(lanes T)**. **Upper part:** a set of Northern blots identical to those presented in Figure 1, lanes 6 and 7, were probed with different subgenomic fragments of cloned WHV DNA. Hybridization and autoradiography were performed as in Figure 1. **Lower part:** location of WHV fragments, used as hybridization probes, on the physical and genetic map of the WHV genome. The different initiation codons of the pre-S and pre-C open reading frames are indicated, and the positions of the stop codons are given by the arrowheads (according to Tiollais *et al.*, 1981 and Galibert *et al.*, 1982).

nique of Southern blotting. After hybridization, uncut DNA from W64 non-tumorous liver and W74 hepatocarcinoma showed an intense band at position 4.2 kb and a long smear underneath (Figure 3A and B, lanes uncut). This demonstrated the presence of actively replicating free viral DNA including the open circular, linear and supercoiled forms (Fowler *et al.*, 1984). After digestion of the cellular DNA by three restriction endonucleases, no band corresponding to high mol. wt. DNA could be detected. However, under our experimental conditions, the presence of small amounts of integrated WHV DNA cannot be excluded. Hybridization experiments of W64 tumorous liver DNA failed to reveal the presence of significant amounts of free viral DNA, whereas discrete bands of high mol. wt. DNA were observed with each endonuclease-digested sample (Figure 3C) suggesting the presence of integrated WHV sequences in cellular DNA.

3′ ends of WHV transcripts

The mapping of the 3′ ends of WHV transcripts was performed using different overlapping probes, covering almost the entire viral genome. Probes *Eco*RI 3047 and *Xba*I 2926 represent the WHV DNA fragments *Pst*I/*Eco*RI* and *Eco*RI/*Xba*I*, respectively (see Figure 2, lower part). They were labelled at their 3′ ends, hybridized to poly(A)$^+$ RNA of W64 and treated with S1 nuclease. The resulting protected fragments indicated a unique 3′ end for all WHV-specific transcripts, around position 2000 within the C gene (data not shown). In order to localize more precisely this 3′ end, probe *Ava*I 942 representing the restriction fragment *Bgl*II/*Ava*I* was annealed to poly(A)$^+$ RNA of W64, total RNA of W65 and total RNA of non-infected W101 (used as a control). A single protected fragment of 464 bases was observed after S1 nuclease and exonuclease VII digestion (Figure 4). Thus, the unique 3′ terminus was localized 20 nucleotides (± 5) downstream of the variant 5′TATAAA3′ polyadenylation signal situated at position 2037 in the WHV genome.

5′ ends of the 2.1-kb mRNA

Northern blot-hybridization experiments clearly demonstrated that the 2.1-kb mRNA covers the region extending from the *Eco*RI site to the *Hind*III site of the WHV genome (Figure 2, lower part, probes I and II). Moreover, an unspliced poly(A)$^+$ RNA species of 2.1 kb, i.e., of $\sim 1.9-2.0$ kb after subtraction of $100-200$ nucleotides for the poly(A) track, which terminates at position 2057 (see above), should initiate a region immediately upstream of the S gene. Thus, we chose the probe *Apa*I 887 and the primer *Apa*I 505, i.e., fragments *Eco*RI/*Apa*I* and *Xba*I/*Apa*I*, respectively, to determine the 5′ end of this mRNA by nuclease mapping and primer extension.

Two fragments of 742 (± 10) bases and 729 (± 10) bases were found to be protected against S1 nuclease as well as against exonuclease VII digestion. The corresponding fragments observed in the primer extension experiment consisted of 738 (± 10) bases and 721 (± 10) bases (Figure 5, A,b). These data localize two efficient transcription initiation sites at positions 145 (± 10) and 158 (± 10) on the WHV genome (see Figures 2 and 5B). In the primer extension experiment, we detected an additional band of 697 (± 10) bases, which had no counterpart in the nuclease mapping pattern. A minor band, corresponding to an extended fragment of 908 (± 20) bases, appeared after longer exposure of the gel (Figure 5, A,a). This fragment is longer than probe *Apa*I 887, and could indicate the presence of a minor transcription start site at nucleotide 3287 (± 20) on the WHV genome, if the transcript is not spliced. We cannot rule out completely the possibility of a splicing event, involving the position 141 as acceptor splice, as proposed by Mandart *et al.* (1984), since we found a strong signal at nucleotide 145 (± 10) in nuclease mapping experiments. Considering positions 145 and 158 as the major start sites and position 2057 as the 3′ end, the calculated length of 1900 nucleotides for the WHsAg RNA would match the length deduced from the Northern blot anlaysis, counting ~ 200 nucleotides for the poly(A) track.

5′ ends of the 3.7-kb mRNA

The second major mRNA of 3.7 kb is slightly larger than the WHV genome (3.3 kb). Since we found only one polyadenylation signal at position 2037 used for the 3′ end processing of all WHV-specific mRNA, this 3.7-kb transcript should start

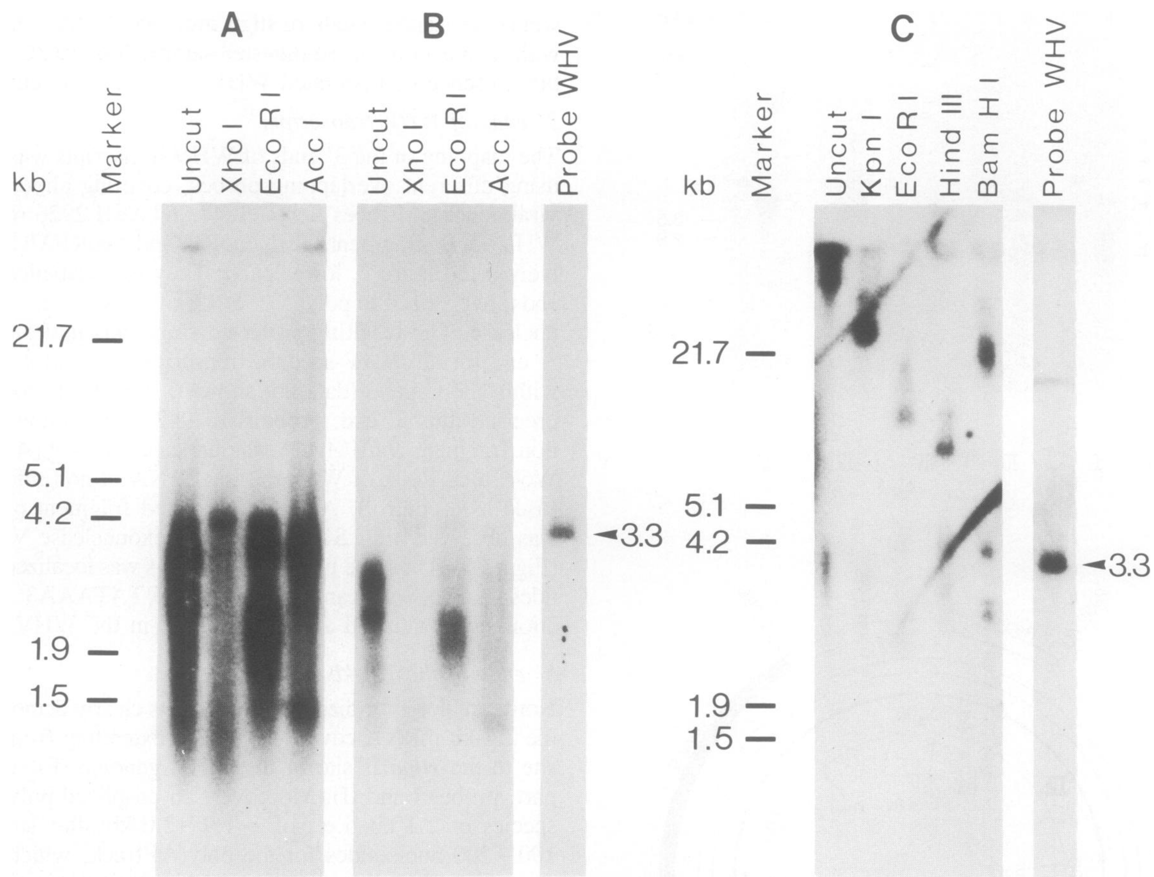

Fig. 3. Southern blotting analysis of liver DNA samples. DNA extracted by SDS/proteinase K digestion, hydrolyzed with different restriction enzymes, analyzed by electrophoresis through 0.8% agarose gels and hybridized with a WHV-specific probe as in Figure 1. The DNA size markers were *Hin*dIII- and *Eco*RI-digested bacteriophage λ DNA and the WHV-specific 3.3-kb *Eco*RI fragment of pBH20-WHV-1, which indicates the position of linear forms of free viral DNA. **(A)** 12 μg of DNA from a non-tumorous liver sample of woodchuck W64 were loaded in each lane. **(B)** 2.5 μg of DNA from hepatocarcinoma tissue of woodchuck W74 per lane. **(C)** 80 μg of DNA from hepatocarcinoma tissue of woodchuck W64 per lane. The blots were exposed for 12-48 h with an intensifying screen at −70°C.

around positions 1800 – 1900. Therefore, we screened the WHV genome for transcription start sites in this region using the probe *Bgl*II 2152 (data not shown) and the probe *Bgl*II 942 (fragments *Xba*I/*Bgl*II* and *Ava*I/*Bgl*II*, respectively). We observed a strong signal at 583 (±10) bases and a less intense signal at 605 (±10) bases by nuclease mapping (Figure 5, A,c). Supporting these data, extended fragments of 587 (±10) bases and 613 (±10) bases were found using primer *Bgl*II 344 (fragment *Hin*dIII/*Bgl*II*) in a primer extension experiment (Figure 5, A,c). The transcription initiation sites corresponding to these protected or extended fragments map at position 1950 (±10) and 1925 (±10) on the WHV genome (see Figures 2 and 5B). In addition, a third extended fragment of 826 (±20) bases could represent a third minor transcription signal at position 1708 (±20), but no corresponding protected fragment could be identified by nuclease mapping (Figure 5). The major start sites correspond to an unspliced transcript of 3420 nucleotides, which terminates at the polyadenylation signal within the core gene. This is in agreement with the size of 3.7 kb estimated for the polyadenylated RNA species from blot hybridization experiments.

Discussion

WHV transcripts were characterized in the liver of chronically infected animals by Northern blot, nuclease mapping and primer extension studies. In livers supporting viral multiplication, the WHV transcripts represented ~0.2% of the total poly(A)$^+$ RNA. Two predominant RNA species of 2.1 kb and 3.7 kb were detected in nearly equal amounts.

The 2.1-kb transcript contains the sequences of the S gene and the X gene, but not of the C gene. Its major initiation sites were mapped at positions 145 and 158 on the viral genome, and were found to terminate at nucleotide 2057. Examination of the WHV sequence reveals that the 5′ start of this RNA is located between the second and third AUG of the S region. Therefore, it can only direct the synthesis of the S gene translation product, initiated at the AUG at position 296 (Figure 2). This polypeptide is 222 amino acids long and corresponds to the major protein of WHsAg, present in the viral envelope in glycosylated and non-glycosylated forms with apparent mol. wts. of 25 and 22 kd (Feitelson *et al.*, 1983). The 39-kd protein of the viral envelope is probably initiated at the second AUG (position 116) of the S region, and could be the translation product of the minor WHV mRNA, initiated at position 3287 on the viral map, 137 nucleotides upstream of the second AUG of the S region (Figure 5). This glycosylated 39-kd protein would be 287 amino acids long. A longer transcript of the S region, initiated upstream of the pre-S open reading frame, which could direct the synthesis of the large protein of the envelope, was not detected in Northern blot experiments. This large protein would also be a translation product of the 3.7 kb RNA.

RNA transcripts similar to the 2.1-kb WHsAg mRNA were described for other viruses of the hepadna group. In the case of duck hepatitis B virus, two different mRNA of 2.1 and 2.35 kb

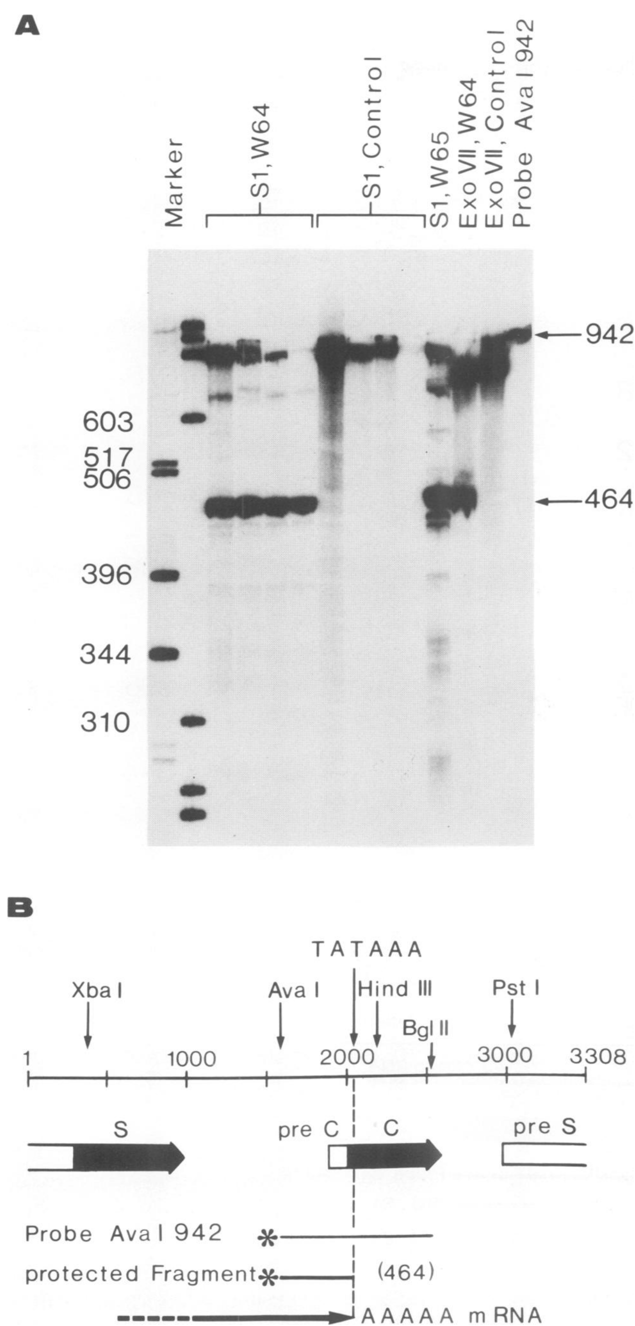

Fig. 4. Nuclease digestion mapping of the 3′ ends of WHV-related transcripts. **(A)** Autoradiogram showing the analysis of WHV RNA 3′ ends by S1 nuclease and exonuclease VII. S1 nuclease assays were performed either with 10 μg of W64 woodchuck poly(A)$^+$, using 500, 1000, 1500 and 3000 units/ml of S1 nuclease (from left to right), or with 40 μg of W65 woodchuck total RNA (1500 units/ml of S1 nuclease). 10 μg of W64 poly(A)$^+$ RNA were analyzed by exonuclease VII mapping using 20 units/ml of the enzyme. The probe *Ava*I 942 labelled at its 3′ end of the minus-strand was used in all experiments. The positions of size markers are give at the right of the lanes. **(B)** Schematic localization of the WHV-specific RNA 3′ processing site. The upper part of the figure represents the physical and genetic map of WHV genome, linearized at its unique *Eco*RI site. The bottom part depicts the location of probe *Ava*I 942, of the protected fragment and of the WHV mRNA processing/polyadenylation site. The labelled 3′ end of probe *Ava*I 942 is indicated by an asterisk.

were characterized (Mason *et al.*, 1982; Büscher *et al.*, 1985). As for WHV, the 2.1 kb RNA initiates downstream of the last AUG of the pre-S region, whereas the longer pre-S mRNA starts at the beginning of the open reading frame, upstream of the second AUG. In the case of HBV, one major 2.1-kb HBsAg mRNA was described (Pourcel *et al.*, 1982; Gough, 1983; Cattaneo *et al.*, 1983; Standring *et al.*, 1984; Michel *et al.*, 1984). Its 5′ end was located a few nucleotides upstream of the second AUG of the pre-S region, suggesting that this transcript could direct the synthesis of the major envelope protein and also of the middle protein, initiated at the second AUG. For this protein, a function of receptor for polymerized human serum albumin was recently proposed (Machida *et al.*, 1983; Michel *et al.*, 1984). Considering the positions of the 5′ ends of WHsAg and DHBsAg mRNA and the short distance between the HBsAg mRNA start site and the initiator codon of the middle protein, it seems more likely that this polypeptide is translated from a minor pre-S mRNA initiated further upstream in the pre-S region (Rall *et al.*, 1983; Cattaneo *et al.*, 1984).

The promoter sequences, involved in positioning the 5′ ends of WHsAg mRNA, differ from 'ATA'-like consensus sequences, as is the case for the other hepadna viruses already studied (Cattaneo *et al.*, 1983; Büscher *et al.*, 1985). It seems probable that a region related to SV40 late promoter sequences is involved in the S gene transcription of WHV as well as of HBV. The variant 5′TATAAA3′ polyadenylation signal, situated within the core gene, is used for WHsAg mRNA termination. An identical variant signal was found in the corresponding region of mammalian hepadna viruses (Galibert *et al.*, 1982; Seeger *et al.*, 1984) and the 3′ end of the HBsAg mRNA was also mapped at this site (Cattaneo *et al.*, 1983; Simonsen and Levinson, 1983).

The second major viral mRNA, 3.7 kb long, covers more than one WHV genome. Two major initiation sites, located at positions 1925 and 1950 on the viral map, were assigned to this transcript, whereas its 3′ processing was found to occur at nucleotide 2057, using the same polyadenylation signal as the 2.1-kb RNA and all WHV transcripts. Therefore, the 3.7-kb RNA is designated as the messenger specifying the viral core protein, using the AUG at position 2021 as the initiation codon for the synthesis of the WHcAg protein, which has the deduced length of 188 amino acids. The sequence preceding the C gene in the same open reading frame is termed here 'pre-C region' and contains two initiation codons, at positions 1910 and 1931 (Figure 2). WHcAg-related polypeptides initiated at these positions could represent the translation products of the minor transcript starting around nucleotide 1708. However, the WHcAg proteins have still to be characterized, and in particular their amino termini have to be defined. It seems likely that 'ATA'-like promoter sequences, situated in the AT-rich region at an appropriate distance upstream of the c-mRNA start sites, are involved in the control of C gene transcription. Similar sequences were found in the corresponding region of DHBV (Büscher *et al.*, 1985). The 3.7-kb mRNA shares with all WHV transcripts a unique polyadenylation site, situated in the C gene, which implies that it is processed only after a second passage through this polyadenylation signal. Furthermore, the minor 5.6- and 6.9-kb RNA species result probably also from multiple passages around the circular WHV genome. The presence of a unique polyadenylation signal, which is used for 3′ processing of all viral transcripts and which is not always efficiently recognized is an unusual feature, common to hepadna viridae.

Transcripts comparable with the 3.7-kb RNA were also found in livers of HBV-infected chimpanzees and DHBV-infected Peking ducks (Cattaneo *et al.*, 1984; Mason *et al.*, 1982; Büscher *et al.*, 1985). In addition, a similar RNA species was identified in a rat cell line transfected with four head-to-tail repeats of HBV DNA (Gough, 1983). The presence of such transcripts was only

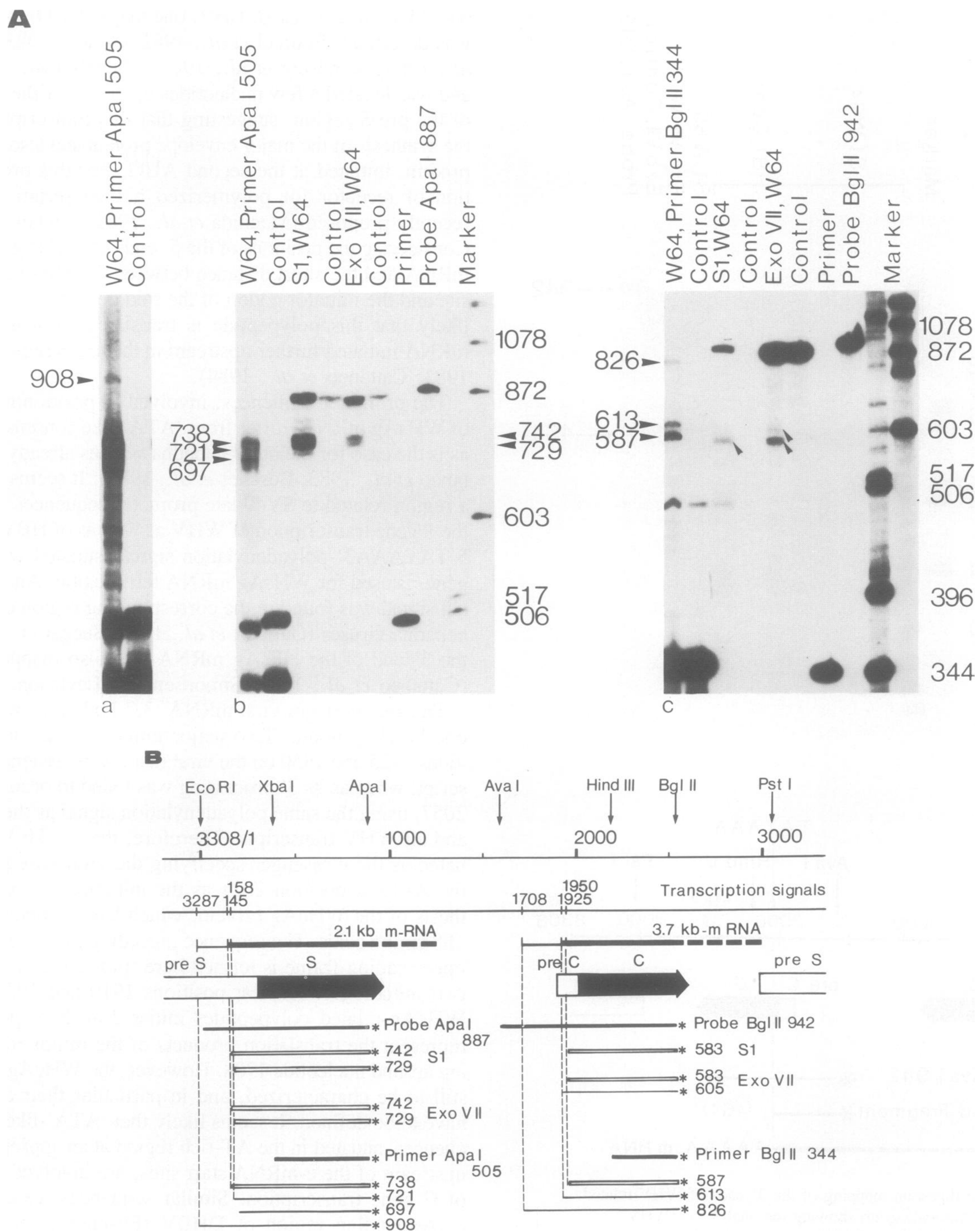

Fig. 5. Mapping of the 5′ ends of the WHV-specific transcripts by S1 nuclease and exonuclease VII digestion and primer extension. **(A)** Products of nuclease digestion and primer extension analysis were separated on 6% polyacrylamide-urea gels and exposed to Kodak XR-5 films at −70°C for 5 h **(b)** or 45 h **(a,c)**. 5−10 μg poly(A)$^+$ RNA of woodchuck W64 were used for nuclease mapping assays and 20 μg for primer extension. Negative controls for the primer extension analysis consisted of 20 μg tRNA. **(B)** Schematic illustration of the location of the transcription signals on the physical and genetic map of the WHV genome. From top to bottom: physical map of the linearized WHV genome, alignment of the transcription signals, the two major mRNA (2.1 kb and 3.7 kb), genetic map of the WHV genome, structure of probes *Apa*I 887 and *Bgl*II 942 and their corresponding fragments, structure of primers *Apa*I 505 and *Bgl*II 344 and their extended fragments. The labelled 5′ end of each probe or primer is indicated with an asterisk.

observed when the virus was replicating or − at a minor level − when the core antigen was expressed.

In the genome of hepadna viruses, two identical 10−12 bp sequences (DR1 and DR2) are directly repeated near the 5′ ends of the viral plus and minus DNA strands (Seeger *et al.*, 1984; Dejean *et al.*, 1984). It is notable that the start sites of the 3.7-kb RNA map within or directly upstream of DR2 (on the plus strand), and that a minor initiation site maps near DR1. In the case of DHBV, the 5′ start of the greater than genome length transcript was also localized a few nucleotides upstream of DR2. The size and the structure of the major WHV 3.7-kb transcript designate this RNA as a candidate for the template of reverse transcription in the pathway of minus-strand DNA synthesis. The terminal redundancies of this molecule, of 100−130 nucleotides,

could play a role in maintaining a circular structure for the intermediate DNA-RNA hybrids, during viral replication (Summers and Mason 1982; Miller *et al.*, 1984). Subsequent degradation of RNA molecules in the hybrids, or during the pathway of plus-strand DNA synthesis, would lead to poly(A)$^-$ intermediates of various lengths, which were observed in Northern blot experiments.

In the absence of any detectable transcription start upstream of the P and X open reading frames, the 3.7-kb and the 2.1-kb RNA could represent suitable messages, using internal initiation codons during translation on polyribosomes. Since no splice has been identified, the message of the P open reading frame could also be translated by frameshift suppression as it is probably the case for an avian retrovirus, or by a mechanism which includes a posttranslational processing.

In the particular case of a HCC liver sample without viral multiplication, the WHV-specific transcripts were ~ 10 times less abundant than in the non-tumorous liver of the same animal, which supported active viral replication. Two poly(A)$^+$ RNA of 4.6 and 2.3 kb were detected in nearly equal amounts, and were shown to contain the sequences of the S and the X gene, but not of the C gene. These RNA species could be transcribed from WHV sequences integrated into cellular DNA or from rearranged episomal forms of high mol. wt. (Rogler and Summers, 1982). From previous studies (Ponzetto *et al.*, 1984) and from the present hybridization data, it seems unlikely that these RNA species could direct the expression of any viral gene other than the S or the X gene. The longer transcript (4.6 kb), could consist of WHV-specific and adjacent cellular sequences and represent a message for a tumor-specific protein. The 3.7-kb RNA absent in this tumor sample, was detected in all liver tissues which supported viral multiplication, reinforcing the hypothesis that this transcript represents the so-called 'pre-genome', acting as a putative template for reverse transcription in the synthesis of viral minus-strand DNA.

Materials and methods

Animals

Woodchucks (*Marmota monax*) were maintained in a colony housed at the Unité de Recherches sur les Hépatites, Lyon, France. The first animal used in this study, W64, had a chronic active hepatitis and produced high levels of WHV virions until February 1984, when a liver biopsy sample was studied. A deterioration in the animal's condition was accompanied by a drop of virus production and a large tumor (~ 15 g) was discovered at autopsy, in July 1984. Tumorous and non-tumorous parts of the liver were resected independently. A second animal, W65, was chronically infected and highly viremic without developing any tumor. Two other chronically infected woodchucks, W74 and W78, that had detectable levels of viral DNA in the serum, were found to have developed a hepatocellular carcinoma during routine liver biopsy. Tumorous and non-tumorous samples were resected separately. The titers of WHV particles present in woodchuck sera were determined by endogenous DNA polymerase activity assay (Summers *et al.*, 1978) as well as by dot hybridization assays, using cloned WHV DNA as a probe (data not shown).

Extraction of liver DNA and Southern blot analysis

Liver tissue samples from surgical biopsies were immediately frozen and kept in liquid nitrogen until used. DNA was prepared from up to 500 mg frozen tissue as previously described (Bréchot *et al.*, 1980). Digestion of liver DNA with different restriction enzymes, conditions of agarose gel electrophoresis, transfer to nitrocellulose and hybridization with [^{32}P]nick-translated WHV DNA were as reported previously (Wahl *et al.*, 1979).

Liver RNA preparation and analysis

Total RNA was extracted from up to 1 g frozen tissue by hot phenol. The tissue was pulverized in liquid nitrogen and transferred into a mixture (1:1) of phenol and lysis buffer (10 mM CH_3COONa, pH 5.2, 50 mM NaCl, 0.5% SDS). After 20 strokes in a Dounce homogenizer, the mixture was incubated for 5 min at 55°C with vigorous agitation and quickly cooled in ice. The aqueous phase was re-extracted twice with phenol and precipitated with ethanol. Polyadenylated RNA was selected by chromatography on an oligo(dT)-cellulose column. Denatured poly(A)$^+$, poly(A)$^-$, or total RNA was separated on 1.2% agarose gels containing 2.2 M formaldehyde and subsequently transferred to nitrocellulose or Zeta-bind (Microfiltration Research Division) filters. Pre-hybridization and hybridization steps were carried out according to standard procedures (Maniatis *et al.*, 1982).

WHV DNA probes

Hybridization probes were prepared either from pBH20-WHV-1, a plasmid containing the WHV genome cloned at its unique *Eco*RI site in pBH20 (Ogston *et al.*, 1982), or from subgenomic clones in pBR322 from pHB20-WHV-1 restriction fragments. Three plasmids termed pWCA, pWCB and pWCC were constructed using, respectively, the *Eco*RI-*Bam*HI fragment (1 – 1529), *Bam*HI-*Bgl*II fragment (1529 – 2535) and *Bgl*II-*Bam*HI fragment (2535 – 3520) (M.Lange, unpublished results). Five different WHV sequences, covering the whole WHV genome, were prepared from these plasmids by digestion with appropriate restriction enzymes, agarose gel electrophoresis, and electroelution. They were radiolabelled by nick-translation to a specific activity of 2 – 4 x 10^8 c.p.m./μg.

The probes used for nuclease mapping and primer extension were derived from the plasmid pBH20-WHV-1. Probes *Bgl*II 942, *Bgl*II 2152 and primer *Bgl*II 344 were labelled at their 5′ ends of the minus-strand to a specific activity of 3 – 5 x 10^5 c.p.m./pmol end. Probe *Apa*I 887 and primer *Apa*I 505 were labelled in the same way to a specific activity of 2 x 10^5 c.p.m./pmol end. Probes *Ava*I 942, *Eco*RI 3047 and *Xba*I 2926 were labelled at the 3′ end of the minus-strand to a specific activity of ~0.8 – 2.8 x 10^6 c.p.m./pmol end.

Nuclease mapping and primer extension

Polyadenylated RNA from woodchuck W64 liver (non-tumorous tissue) and total RNA from woodchuck W65 liver were used. In the latter case, the method of Auffray and Rougeon (1980) was applied for RNA preparation. Negative controls consisted of total RNA of non-infected woodchuck W101 liver, prepared as described above.

S1 nuclease and exonuclease VII assays were performed with minor modifications as described previously (Weaver and Weissman, 1979; Berk and Sharp, 1978; Favoloro *et al.*, 1980). Briefly, either 20 – 30 μg of total cellular RNA, or 5 – 10 μg of poly(A)$^+$ RNA were incubated with 0.1 – 0.15 pmol of probe *Bgl*II 942 or 0.2 – 0.3 pmol of probe *Apa*I 887. All samples were brought to a total of 40 μg RNA with *Escherichia coli* rRNA and then hybridized in 80% formamide at 51°C for at least 9 h. The hybrids were digested with 500 – 4000 units/ml S1 nuclease (Sigma Type III) for 1 h at 22°C or with 20 – 40 units/ml exonuclease VII (BRL) at 37°C for 2.5 h. After ethanol precipitation, the protected fragments were denatured by heating in 80% formamide and loaded on a 6% polyacrylamide-8 M urea gel.

For primer extension, the hybrids were passed through a Biogel column (1 ml), precipitated and washed with 70% ethanol. The reaction was carried out at 42°C for 1 h in the presence of four unlabelled deoxynucleotides, using 50 units of reverse transcriptase from avian myeloblastosis virus (AMV, Life Sciences). After phenol extraction, the samples were precipitated, washed with 70% ethanol, denatured by heating in 80% formamide and loaded on 6% polyacrylamide-8 M urea gels.

Acknowledgements

We thank C.Pichoud and C.Jacquet for providing woodchuck liver samples, Y. Malpièce, M.Kaczorek and E.Dassa for technical advice and helpful discussions, and A.Cova and L.M.Da for typing the manuscript. T.M. received support from the Commission of the European Communities. This work was supported by grants from the National Institutes of Health and from the Association pour la Recherche sur le Cancer.

References

Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303-314.

Berk,A. and Sharp,P.A. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 1274-1278.

Bréchot,C., Pourcel,C., Louise,A., Rain,B., and Tiollais,P. (1980) *Nature*, **286**, 533-535.

Büscher,M., Reiser,W., Will,H. and Schaller,H. (1985) *Cell*, in press.

Cattaneo,R., Will,M., Hernandez,N. and Schaller,M. (1983) *Nature*, **305**, 336-338.

Cattaneo,R., Will,M. and Schaller,M. (1984) *EMBO J.*, **3**, 2191-2196.

Chakraborty,P., Ruiz-Opalo,N., Shouval,D. and Shafritz,D. (1980) *Nature*, **286**, 531-533.

Dejean,A., Sonigo,P., Wain-Hobson,S. and Tiollais,P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5350-5354.

Edman,J., Gray,P., Valenzuela,P., Rall,L. and Rutter,W.J. (1980) *Nature*, **286**, 535-538.

Favoloro,J., Treisman,R. and Kamen,R. (1980) *Methods Enzymol.*, **65**, 718-749.

Feitelson,M., Marion,P. and Robinson,W. (1981) *J. Virol.*, **39**, 447-454.

Feitelson,M., Marion,P. and Robinson,W. (1983) *Virology*, **130**, 76-90.

Fowler,M.J.F.,Monjardino,J., Tsiquaye,K.N., Zuckermann,A.J. and Thomas, H.C. (1984) *J. Med. Virol.*, **13**, 83-91.
Frommel,D., Crevat,D., Vitvitsky,L., Pichoud,C., Hantz,O., Chevalier,M., Grimaud,J.A., Lindberg,J. and Trepo,C. (1984) *Am. J. Pathol.*, **115**, 125-134.
Galibert,F., Chen,T. and Mandart,E. (1982) *J. Virol.*, **39**, 447-454.
Gerin,J.L. (1984) in Chisari,F. (ed.), *Advances in Hepatitis Research*, Masson Publishing USA Inc., pp. 40-48.
Gough,N. (1983) *J. Mol. Biol.*, **165**, 683-699.
Gough,N. and Murray,K. (1982) *J. Mol. Biol.*, **162**, 43-68.
Laub,O., Rall,L., Fruett,M., Shaul,Y, Standring,D., Valenzuela,P. and Rutter, W.Y. (1983) *J. Virol.*, **48**, 271-280.
Machida,A., Kishimoto,S., Ohnuma,H., Miyamoto,H., Baba,K., Oda,K., Nakamura,T., Miyakawa,Y. and Mayumi,M. (1983) *Gastroenterology*, **85**, 268-274.
Mandart,E., Kay,A. and Galibert,F. (1984) *J. Virol.*, **49**, 782-792.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Marion,P., Oshiro,L., Regnery,D., Scullard,G. and Robinson,W. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 2941-2945.
Mason,W., Seal,G. and Summers,J. (1980) *J. Virol.*, **36**, 829-836.
Mason,W.S., Taylor,J.M., Seal,G. and Summers,J. (1982) in Szumness,W., Alter,M.J. and Maynard,J.E. (eds.), *Viral Hepatitis*, Franklin Institute Press, Philadelphia, pp. 107-116.
Michel,M.L., Pontisso,P., Sobzack,E., Malpièce,Y., Streeck,R.E. and Tiollais, P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7708-7712.
Miller,R.H., Tran,C.T. and Robinson,W.S. (1984) *Virology*, **139**, 53-63.
Ogston,C.W., Jonak,G.J., Rogler,C.E., Astrin,S.M. and Summers,J. (1982) *Cell*, **29**, 385-394.
Ponzetto,A., Cote,P., Ford,E., Purcell,R. and Gerin,J. (1984) *J. Virol.*, **52**, 70-76.
Pourcel,C., Louise,A., Gervais,M., Chenciner,N., Dubois,M.F. and Tiollais,P. (1982) *J. Virol.*, **42**, 100-105.
Rall,L., Standring,D., Laub,O. and Rutter,W.Y. (1983) *Mol. Cell. Biol.*, **3**, 1766-1773.
Robinson,W.S., Marion,P., Feitelson,M. and Siddiqui,A. (1982) in Vyas,G.N., Cohen,S.N. and Schmid,R. (eds.), *Viral Hepatitis*, Franklin Institute Press, Philadelphia, pp. 57-68.
Rogler,C.E. and Summers,J. (1982) *J. Virol.*, **44**, 852-863.
Seeger,C., Ganem,D. and Varmus,E. (1984) *J. Virol.*, **51**, 367-375.
Simonsen,C. and Levinson,A.D. (1983) *Mol. Cell. Biol.*, **3**, 2250-2258.
Snyder,R. and Summers,J. (1980) in *Cold Spring Harbor Conference of Cell Proliferation VII, Viruses in Naturally Occurring Tumors*, Cold Spring Harbor Laboratory Press, NY, pp. 447-457.
Standring,D., Rutter,W., Varmus,H. and Ganem,D. (1984) *J. Virol.*, **50**, 563-571.
Summers,J. (1981) *Hepatology*, **1**, 179-183.
Summers,J. and Mason,W.S. (1982) *Cell*, **29**, 403-415.
Summers,J., Smolec,J. and Snyder,R. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4533-4537.
Tiollais,P., Charnay,P. and Vyas,G.N. (1981) *Science (Wash.)*, **213**, 406-411.
Tiollais,P., Dejean,A., Brechot,C., Michel,M.L., Sonigo,P. and Wain-Hobson, S. (1984) in Vyas,G.N., Dienstag,J.L. and Hoofnagle,J.M. (eds.), *Viral Hepatitis and Liver Diseases*, Grune and Stratton, Inc., NY, pp. 49-65.
Wahl,G.M., Stern,M. and Stark,G.R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3683-3687.
Weaver,R. and Weissman,C. (1979) *Nucleic Acids Res.*, **7**, 1175-1193.
Will,H., Cattaneo,R., Pfaff,E., Kuhn,C., Roggendorf,M. and Schaller,H. (1984) *J. Virol.*, **50**, 335-342.

Received on 1 March 1985; revised on 12 April 1985