The EMBO Journal vol.7 no.13 pp.4185-4192, 1988

The amino-terminal domain of the hepadnaviral P-gene encodes the terminal protein (genome-linked protein) believed to prime reverse transcription

Ralf Bartenschlager and Heinz Schaller

Zentrum für Molekulare Biologie Heidelberg (ZMBH), University of Heidelberg, Im Neuenheimer Feld 282, 6900 Heidelberg, FRG

Communicated by H.Schaller

A series of antisera directed against amino acid sequences from different segments of the duck hepatitis B virus (DHBV) *P*-gene were shown to immunoprecipitate DHBV DNA molecules that were covalently linked to the DHBV DNA terminal protein. Restriction analysis and sizing after protease treatment demonstrated that the *P*-gene proteins were bound to the 5'-end of the DHBV DNA minus-strand which was mapped to a G-residue in the centre of the repeat sequence DR1. Resistance to alkali treatment indicated a phosphodiester linkage to tyrosine between protein and DNA. Limited protease treatment prior to immunoprecipitation cleaved C-terminal P-proteins from the viral DNA, indicating that the terminal protein forms a separate domain encoded in the N-terminal part of the *P*-gene. Functional analysis of a deletion mutant confirmed the notion that a non-essential spacer separates the terminal protein from the polymerase domain residing in the C-terminal half of the *P*-gene. Thus, the major proteins required for hepadnaviral reverse transcription, namely the primer, DNA polymerase, and possibly also RNase H, appear to be synthesized as a polyprotein precursor which is at least initially linked as such to its first DNA product.

Key words: duck hepatitis B virus/DNA terminal protein/protein primer

Introduction

Hepatitis B viruses, also called hepadna viruses, form a family of small enveloped animal viruses that are characterized by a distinct liver tropism and a narrow host range. These viruses contain a small, partially double-stranded DNA genome which is held in a circular conformation by base-pairing between the overlapping 5'-ends of the DNA strands. Despite containing a DNA genome, hepadna viruses are related to the RNA-containing retroviruses by a common replication mechanism, which for both virus families, involves the synthesis of an unspliced, terminally redundant genomic RNA that is packaged into a viral nucleocapsid and reverse transcribed into a DNA genome. However, distinct differences exist: in HBV replication the onset of reverse transcription occurs before virus maturation while in retroviral replication it is delayed until the mature virus re-enters a host cell. This explains why retroviruses contain an RNA genome, hepadna viruses an uncompleted double-stranded DNA genome. Furthermore, the hepadnaviral genome contains a genome-linked protein attached to the 5'-end of its minus strand, suggesting that hepadnaviral reverse transcription is primed by a protein and not by a tRNA molecule as is the case in retroviral replication (for a recent review, see Ganem and Varmus, 1987).

Protein-primed reverse transcription has not yet been observed in any other replication system; related examples are only known for viral DNA→DNA replication or RNA→RNA replication (e.g. in adenoviruses or in picornaviruses, respectively). The mechanism of this unconventional process is therefore of general interest, especially since it remains unclear whether the hepadnaviral genomic protein is virally encoded or of cellular origin.

In principle, this question should readily be answered since hepadnaviral genomes are very small (3.0–3.2 kb) and contain only three (in mammalian hepadna viruses, four) open reading frames (ORFs) encoding viral proteins. Of these, two have been assigned to structural genes, *preS/S* and *C*, whose well-characterized products form the major constituents of the viral envelope and the nucleocapsid, respectively. The only other gene present in all hepadna viruses is the *P*-gene which is believed to encode the viral DNA polymerase/reverse transcriptase as indicated by sequence homologies in its 3'-proximal sequence to retroviral *pol* genes (Toh *et al.*, 1983), and by mutational analysis of the *P*-ORF in the HBV and DHBV system (Radziwill, 1988; Schlicht *et al.*, 1988). A fourth ORF, the *X*-gene, which appears to encode a transcriptional transactivator (Spandau and Lee, 1988; Zahm *et al.*, 1988), is absent from the genome of avian hepadna viruses; it therefore cannot encode the terminal protein found in all hepadnaviral genomes. This leaves a *P*-frame product as the only possible candidate for a virally encoded genome-linked protein.

Unfortunately, the structural and biochemical properties of the hepadnaviral genomic protein remain largely unknown. Although recognized quite early in the case of HBV due to its property of carrying the viral DNA into a phenol phase, a thorough biochemical characterization of the HBV protein has been hampered by its low abundance (possibly only a single molecule per virion), and also by properties unfavourable for protein purification and characterization as noted in the early work of Gerlich and Robinson (1980). Most later work has therefore been performed using the duck hepatitis B virus (DHBV) as an animal model which provides access to larger amounts of virus and virus-infected tissue. In this system, the DNA components of protein-linked replication intermediates have been well characterized (Molnar-Kimber *et al.*, 1983; Lien *et al.*, 1987). Furthermore, systems for the functional analysis of mutated viral genomes *in vivo* and in tissue culture have been established (Sprengel *et al.*, 1985; Schlicht *et al.*, 1987b). For these reasons, we have chosen the DHBV system to investigate the origin of the hepadnaviral genomic protein. The results obtained provide direct experimental evidence that this protein is encoded in the viral *P*-gene and that it forms a structural domain separate from that of the DNA polymerase/reverse transcriptase. By analogy to the

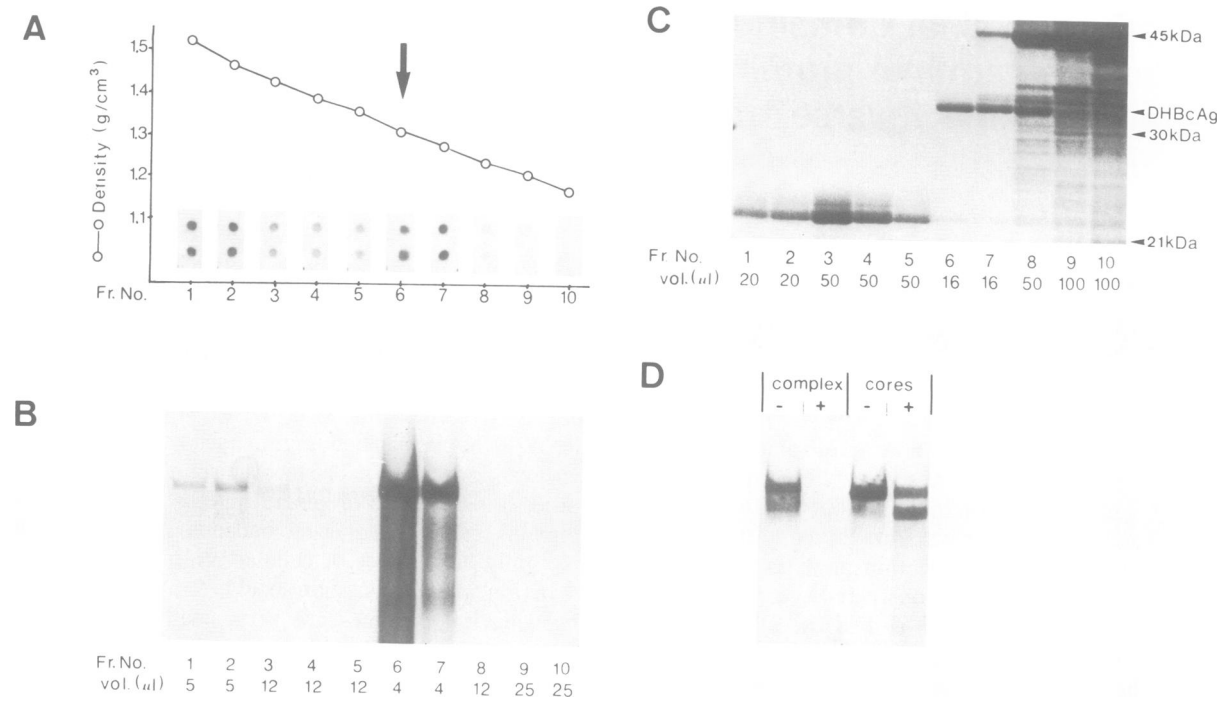

Fig. 1. Purification of DHBV nucleocapsids and of a non-encapsidated DHBV DNA−protein complex, capable of endogenous DNA synthesis. From the final CsCl-gradient (see Materials and methods), 25 fractions were collected from the bottom of the tube and the first 10 analysed in detail. **(A)** Measurement of DHBV DNA by dot blot analysis. 1 µl of each fraction was spotted on a nitrocellulose filter and hybridized with a ^{32}P-labelled, nick-translated probe. The arrow indicates the position of the peak of radiolabelled nucleocapsids, added as 'marker particles' at the beginning of the purification. **(B)** Aliquots of each fraction corresponding to 10 ng DHBV-DNA were assayed for endogenous DNA polymerase activity and the ^{32}P-labelled DNA was analysed on an SDS-containing 1.2% agarose gel. **(C)** Analysis of proteins contained in aliquots of each fraction on a 12.5% discontinuous polyacrylamide gel. The positions of core protein (DHBcAg) and mol. wt markers are shown on the right. **(D)** Nuclease sensitivity of viral DNA from core particles and the DNA−protein complex. Viral DNA from pools of fractions 1 and 2 (complex) and 6 and 7 (cores) was radiolabelled in the endogenous polymerase assay and analysed on an SDS-containing agarose gel before (−) and after (+) treatment with micrococcal nuclease.

terminology used in the protein-primed DNA replication of adenoviruses, this protein is referred to as the 'DHBV DNA terminal protein' (TP) throughout this report.

Results

Purification of core particles and isolation of a non-encapsidated replication complex

To obtain the DHBV DNA TP in sufficient quantity and largely free from contaminating cellular components, a combination of pre-existing methods was adopted for the purification of viral nucleocapsids from large amounts of DHBV-infected duck liver (Bartenschlager, 1987; Materials and methods). As judged from DNA dot blot analysis, this method yielded ~100 pmol per kg of duck liver of nucleic acid containing DHBV cores (corresponding to ~10% of the viral DNA present in the liver tissue) and presumably also an equal amount of terminal protein.

In addition to this core fraction, an equal amount of DHBV DNA (and genomic protein) was recovered in a second DHBV DNA-containing fraction in the last purification step, a CsCl-density gradient. As shown in Figure 1, this fraction sedimented to a much higher density (1.45−1.5 in the experiment shown in Figure 1A, but 1.6 in equilibrium runs) than that of the normal nucleocapsids (1.3). It also differed from the core fraction in that it contained, normalized to DHBV DNA, much less core protein (Figure 1C) and a reduced enzymatic activity in the endogenous polymerase reaction (Figure 1B). Furthermore, the DNA contained in this fraction was highly sensitive to digestion with micrococcal nuclease whereas the DNA in the core fraction was not (Figure 1D). However, the viral DNA was equally extractable into a phenol layer (not shown). Taken together, these results suggest that this 'heavy' DNA−protein complex consists of viral genomes, at least partially enzymatically active DNA polymerase, and terminal protein, which had largely lost their surrounding shell of core protein upon exposure to high salt. Although still containing substantial amounts of a 23 kd duck liver protein of unknown identity, this fraction was used as a preferred source for further analysis of the DHBV DNA-associated P-proteins, since the core protein unspecifically absorbs to nucleic acids (see below).

Anti-P antisera immunoprecipitate protein linked to the viral DNA minus-strand

A panel of antisera (Figure 2A), raised against *P*-ORF fusion proteins or synthetic peptides and shown to be capable of immunoprecipitating DBHV *P*-gene products synthesized in an *in vitro* translation system (Bosch *et al.*, 1988), was used to identify *P*-proteins in the DHBV DNA−protein complex. Immunoprecipitations were monitored in a subsequent dot blot analysis using a single stranded DHBV DNA probe of plus-strand polarity. This assay also allows to test whether P-proteins remain linked to the viral DNA minus-strand upon exposure to conditions that remove non-covalently linked proteins (e.g. boiling in 2% SDS, 2% 2-mercaptoethanol).

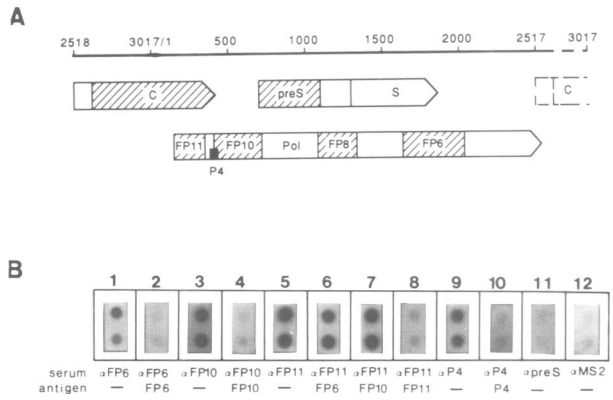

Fig. 2. Immunoprecipitation of protein-linked DHBV DNA minus-strands with P-frame specific antisera. **(A)** Linear representation of the DHBV-genome showing the positions of the viral genes. The regions of these genes expressed in *E.coli* and used to elicit segment-specific antisera are hatched. The position of the peptide P4 is also indicated. **(B)** Detection of immunoprecipitated DNA by dot blot hybridization. Aliquots of the DNA−protein complex were analysed in duplicates, with or without added competing antigens, with the antisera indicated as described in Materials and methods.

Therefore, it enabled us to determine directly the relatedness of the DHBV-TP to DHBV *P*-gene products.

The results of such an experiment are presented in Figure 2. As shown in Figure 2B (lanes 1, 3, 5 and 9), all anti-P antisera tested recognized a protein that is tightly associated with the DHBV genome. Recognition was sequence-specific since the immunoprecipitations were inhibited by the presence of the homologous antigens (lanes 2, 4, 8 and 10) but not by products derived from other parts of the *P*-gene sequence (lanes 6 and 7). Comparable results were also obtained with an anti-peptide antiserum directed against a *P*-gene encoded sequence (P4, lanes 9 and 10) which partially overlaps with fusion protein FP10. Control sera, raised against other DHBV gene products (e.g. the preS surface antigen, lane 11) or the MS2 polymerase leader sequence employed in the bacterial fusion proteins (lane 12) gave negative results.

In addition to the reaction with anti-P antisera, a minor signal was also consistently observed with an anti-core antiserum (αC, Figure 3). However, the immunoprecipitated DHBV DNA fraction was low compared to that recognized by the anti-P antisera and no further signal was obtained in a second immunoprecipitation (Table I). This suggests that the core-protein had not been completely removed from the DNA−protein complex fraction, and that some of it remained non-specifically associated with a minority of the viral genomes even after denaturation by boiling in 2% SDS. This interpretation is supported by the finding that significantly higher anti-core related signals were obtained if we used DHBV nucleocapsids as a starting material instead of the high density complex. Otherwise, results very similar to those presented in Figure 2 were obtained with denatured DHBV cores (not shown).

The DNA minus-strand is linked to a defined protein domain located at the N-terminus of the P-frame product

To examine whether the DHBV DNA minus-strand was linked to a specific part of the P-protein sequence, we asked whether parts of the protein could be preferentially removed from the DNA−protein complex by protease digestion. For this purpose, time course experiments were performed with various proteases, and followed by immunoprecipitation with anti-P antisera specific for different portions of the *P*-gene product and subsequent dot blot analysis with a plus-stranded DNA probe. Figure 3 shows the results of the kinetics of digestion with protease V8 from *Staphylococcus aureus*. It demonstrates that polypeptide sequences recognized by anti-FP6 and anti-FP8 were rapidly cleaved from the DNA−protein complex, whereas those reacting with anti-FP10 or anti-FP11 were largely refractory to V8 action. Of the latter two, the P10 region appeared consistently to be more protease-resistant and was sometimes immunoprecipitated even better during the early stages of protease action (Figure 3 and Table II).

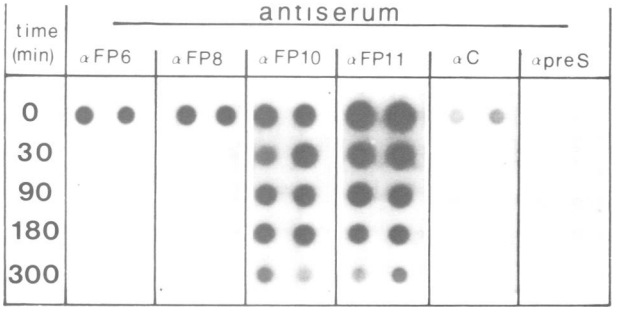

Fig. 3. Kinetics of protease action on different protein domains in the DNA-linked *P*-gene product as detected by immunoprecipitation with segment-specific antisera. The denatured DNA−protein complex, prepared as described in Materials and methods, was digested with protease V8 for the times indicated, and assayed for protein-linked DHBV DNA by immunoprecipitation and subsequent dot blot analysis. A quantitation of the results is given in Table II.

Table I. Percentage of DHBV DNA detected in successive immunoprecipitations[a]

	Antiserum				
	αFP6	αFP10	αFP11	αC	αMS2
1st precipitation	11	17	40	7	2
2nd precipitation[b]	12	9	7	2	1

[a]Percent of signal obtained with input DNA (3100 c.p.m.); experimental procedures as in Figure 2.
[b]Result obtained with the supernatant from the first immunoprecipitation.

Taken together, these results provide strong evidence that the DHBV DNA minus-strand is linked to the N-terminal region of the DHBV *P*-gene product, and herein probably to an amino acid sequence contained in the fusion protein FP10. The rapid release of the more carboxy-terminal P-protein segments suggests, furthermore, that the V8-resistant N-terminal sequence covering the P10 and P11 regions represents a tightly folded structural domain that is separated from the rest of the *P*-gene product by a rather poorly structured, and therefore highly protease-sensitive tether. This interpretation is also supported by earlier sequence comparisons (Sprengel *et al.*, 1985) demonstrating that the N-terminal domain (now identified as the TP domain) is evolutionarily well conserved, whereas the adjacent spacer region, which also encodes in the overlapping +1 frame the preS protein, is not.

Table II. Quantitation of DHBV DNA (in c.p.m.) immunoprecipitated after limited digestion with protease V8[a]

Incubation time (min)	Antiserum							
	αFP6	αFP8	αFP10	αFP11	αC	αpreS	αFP10[b]	αFP11[b]
0	224	252	311	836	100	29	341	1348
30	41	74	477[c]	464	59	21	674	1291
90	16	50	290	427	57	29	647	953
180	22	75	291	243	53	36	540	639
300	23	35	180	145	30	30	460	519

[a]Mean values of the radioactivity (Cerenkov counts) determined in the dots from the experiment shown in Figure 3.
[b]Values from a second experiment (not shown).
[c]Only one value.

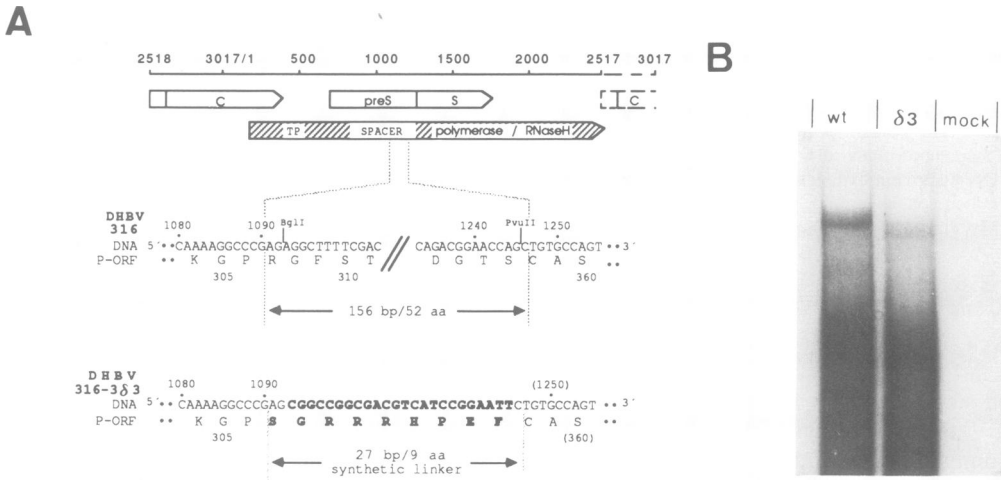

Fig. 4. Functional analysis in cell culture of a DHBV deletion mutant in the *P*-gene spacer region. **(A)** Linear representation of the DHBV genome. Functional domains of the *P*-gene are indicated by shading. Below, sequence alterations in the deletion mutant DHBV 316-3δ3 and the corresponding 316 wildtype sequence are shown (see text). **(B)** Transient expression of these constructs in cell culture. Core particles, isolated from cell lysates by immunoprecipitation were subjected to an endogenous polymerase reaction and nucleic acids analysed on an SDS-containing agarose gel.

To test the prediction of a non-essential P-frame spacer experimentally, a mutant DHBV genome, DHBV 316−3δ3 (Klingmüller, 1988) that carried a large deletion in the *preS/P* spacer region was analysed for its capacity to synthesize functional *P*-frame products. In this mutant, 153 bp from the *preS/P* spacer had been replaced by a synthetic oligonucleotide of 24 bp; as a consequence, 52 amino acids had been deleted in the P frame and replaced by a 9 amino acid linker of different sequence (Figure 4A). After transfection into HepG2 cells, the enzymatic activity of *P*-gene products from this construct was analysed by measuring the *in vitro* incorporation of ^{32}P-labelled deoxynucleotides into immature viral genomes contained in cytoplasmic derived nucleocapsids (endogenous polymerase assay, see Materials and methods). Figure 4B presents a size analysis of labelled DNA products obtained in this assay upon transfection with DNA from the deletion mutant or its parent strain, DHBV 316. They demonstrate that nucleocapsids from the deletion mutant contain normal amounts of DNA polymerase activity, and very similarly labelled, but slightly smaller DNA products than those produced by wild type DHBV. Analogous results were obtained with DHBV mutants carrying other deletions or insertions in the *preS/P* spacer region (Klingmüller, 1988). Thus, despite drastic changes of sequence information in the spacer region, mutant P-frame products are still capable of catalysing the several steps required for viral DNA synthesis: priming and synthesis of the DNA minus-strand, removal of the RNA pregenome, as well as plus-strand DNA synthesis.

The DNA−protein linkage most likely involves a phosphodiester bond between a tyrosine residue and a deoxyguanosine nucleotide

An analysis of the nature of the protein−DNA linkage has so far proven to be quite difficult since the protein component of the complex appears to be heterogenous and possesses a great tendency to aggregate even under routinely used denaturing conditions (Bosch *et al.*, 1988). To achieve better detection and resolution we therefore reduced the size of the DNA-linked polypeptide by extensive protease digestion. In addition, the protein-linked DNA minus-strand was also shortened by cleavage with restriction nucleases. The latter was done after ^{32}P-labelling of nascent DNA minus-strands by carrying out the endogenous reverse transcription reaction using immature core particles capable of early minus-strand synthesis *in vitro* (Summers and Mason, 1982; Materials and methods).

Optimal results were obtained by using trypsin as a protease and *Nco*I as restriction nuclease. This combination yielded a labelled DNA fragment which co-migrated in a sequencing gel with a co-linear 187 nucleotide long marker DNA fragment which differed only at its 5′-end (Figure 5B, lanes 3 and 4). No signal was obtained after treatment with protease V8 (lane 2) indicating that a V8-resistant protein

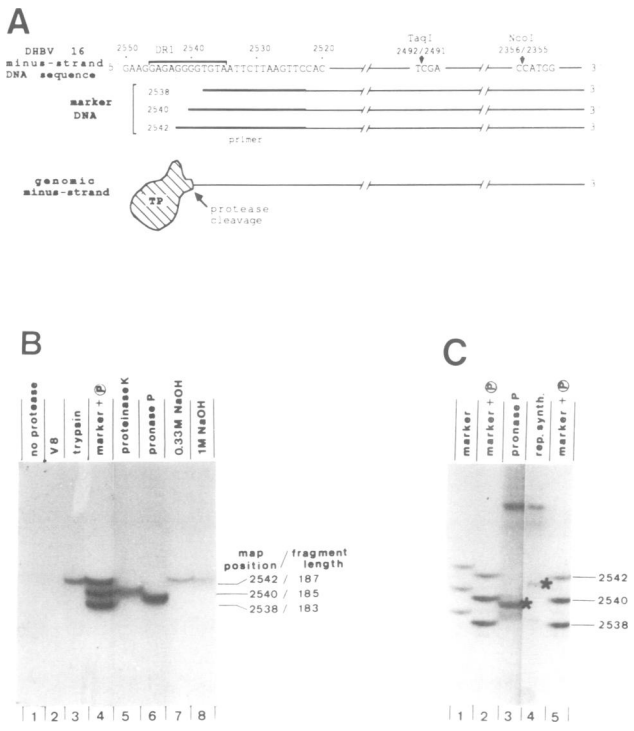

Fig. 5. Resistance of the DNA−TP linkage to alkali cleavage and mapping of the 5′-end of DNA minus-strand. (**A**) Schematic outline of the relevant nucleotide sequences from the DHBV minus-strand near the direct repeat sequence 1 (DR1). Below, three oligonucleotides with 5′-ends at positions 2538, 2540 and 2542 and a common 3′-end at position 2523, which were used to produce DNA markers by primer extension, are shown (solid line). The genomic minus-strand with the TP linked to its 5′-end is shown on the bottom. (**B**) Analysis of the stability of the DNA−TP-linkage to alkali treatment. Nascent DNA minus-strands in core particles were radiolabelled in the endogenous polymerase reaction, treated with the proteases indicated and cut with *Nco*I as described under experimental procedures. To analyse the stability of the DNA−TP-linkage, trypsin digested material was incubated either in 0.33 M NaOH for 2 h at 37°C, or in 1 M NaOH for 6 min at 95°C. Finally, all fragments were ethanol precipitated and analysed on a 6% denaturing polyacrylamide gel. Marker + P = mixture of co-linear 5′-phosphorylated DNA size markers, prepared by primer extension and *Nco*I cleavage as outlined in (A). Numbers on the right refer to the 5′-ends of the oligonucleotides and the fragment lengths. (**C**) To map the 5′-end of DNA minus-strand, a *Taq*I fragment with the pronase P treated TP at its 5′-end was prepared as described under (B) and analysed on a 8% denaturing polyacrylamide gel in parallel with 5′-phosphorylated and unphosphorylated marker DNA fragments, also cleaved with *Taq*I. An aliquot of the product was subjected to DNA repair synthesis with Klenow enzyme and analysed in parallel (lane 4). The genomic minus-strand fragment is marked with a (*).

had carried the DNA into phenol prior to gel-electrophoresis. Digestion with pronase P or proteinase K increased the mobility of the labelled DNA equivalent to a loss of two or three nucleotides, respectively (Figure 5B, lanes 5 and 6), indicating that the DNA-linked tryptic peptide consisted of only very few amino acid residues which were removed to the ultimate or penultimate amino acid by pronase P treatment.

This size increment related to the presence of the peptide allowed us to quantitatively monitor the stability of the DNA−protein linkage and to obtain more information on its chemical nature. As shown in Figure 5B, sizing on DNA sequencing gels was used to examine whether the DNA−peptide bond resisted high pH treatment known to cleave nucleoside phosphodiester bonds to serine or threonine, but not to tyrosine residues. No size shift of the labelled DNA fragment, indicating a loss of the DNA-linked tryptic peptide, was observed after a 2 h incubation in 0.33 M NaOH at 37°C (lane 7) or under much more vigorous conditions (6 min at 95°C in 1 M NaOH, lane 8). By exclusion, it thus seems to be likely that the DHBV DNA TP is linked to the polynucleotide chain via a phosphodiester tyrosine bond. This type of protein−nucleotide linkage has also been observed in the genomes of picornaviruses (Rothberg *et al.*, 1978) and parvoviruses (Chow *et al.*, 1986).

The high resolution on the sequencing gel of the various DNA−peptide species also enabled us to identify with high precision the position of the peptide-linked nucleotide residue relative to the 5′-ends of three closely spaced, co-linear DNA size markers. These markers had 3′-ends identical to the peptide-linked DNA fragment since they had been created in either case by cleavage with *Nco*I at position 2556 or by *Taq*I at position 2492. Cleavage by the latter resulted in shorter fragments (47, 49 and 51 nt long, instead of 183, 185 and 187 with *Nco*I, Figure 5A), and therefore in a higher resolution; this was paralleled by a much reduced labelling of the genomic DNA fragment, and therefore, also in weaker signals relative to the non-specific background on the X-ray film. The results (Figure 5C) demonstrate that the genomic DNA fragment migrated slightly faster than a marker fragment that had a phosphorylated 5′-end at position 2540. (Several other bands on the gel, including a weak band at position 2538.5, appear to be unrelated to the 5′-terminal minus-strand fragment since their migration was, in contrast to the main band, not affected either by the addition of two nucleotides at the *Taq*I cleavage site by repair synthesis (lane 4) or by the size increase of the DNA-linked peptide created by treatment with proteinase K instead of pronase P digestion, not shown.) For an exact evaluation of these data it should be noted that the pronase-treated genomic DNA fragment and the marker fragments differ at their 5′-ends (a phosphodiester-linked tyrosine or dipeptide, or a 5′-phosphate, respectively) only slightly in size and not by charge. These differences most likely contribute less than one nucleotide equivalent to the fragment's electrophoretic mobility on a sequencing gel. Therefore, the data showing that it migrates slightly ahead of a marker starting at nucleotide 2540 (Figure 5C) are taken to indicate that the 5′-end of the DNA minus-strand is located at position 2539.

Discussion

A limited number of virus families are characterized by containing a protein linked to the 5′-end of their DNA or RNA genome (Daubert and Breuning, 1984; Wimmer, 1982). Shown in most cases to be virally encoded, these proteins are generally assumed to function as primers for the replication of the respective viral genomes. Uncertainties as to their viral origin remained for some small genomes of very limited coding capacity, but these have recently been resolved for parvoviruses (Cotmore and Tattersall, 1988), and now, by the present study, also for the hepadna viruses. This adds further evidence supporting the assumption that the hepadnaviral terminal protein functions as a primer for viral DNA replication.

The work described here for DHBV, and as well as similar, but less complete data for HBV (R.Bartenschlager,

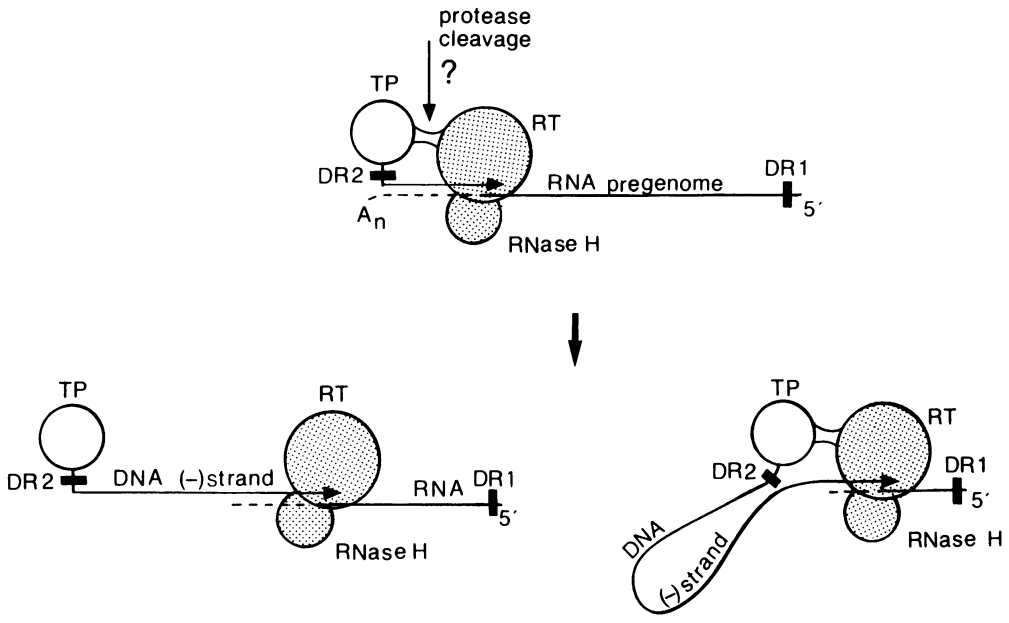

Fig. 6. Scheme of hepadnaviral reverse transcription with or without protease cleavage of the *P*-gene product. TP, terminal protein; RT, reverse transcriptase; DR, direct repeat sequence. The second copy of DR1 that is the site of replication initiation, but only partly copied into DNA (see Figures 5 and 7), is not shown at the TP–DNA minus-strand junction.

unpublished), demonstrate convincingly that the hepadnaviral terminal protein is of viral origin. Together with two other recent reports, our data also offer an explanation for the failure of previous attempts to clarify this point. As shown in Table I, P-proteins are much less immunogenic and antigenic than the structural proteins routinely used as a reference for the sensitivity of immunological assay systems. In addition, hepadnaviral *P*-gene products are tightly linked to their DNA template, appear to be heterogeneous, and form insoluble aggregates (Bosch *et al.*, 1988; Radziwill *et al.*, 1988).

Due to an improved purification procedure for large amounts of viral nucleocapsids and of a non-encapsidated DNA–protein complex, and by using highly sensitive DNA detection techniques, we have finally gained some insight into the structural and functional organization of the *P*-gene and of its products. The biochemical and serological data provided make it evident that the terminal protein forms a distinct domain comprising the amino terminal quarter of the 785 amino acid long multifunctional primary *P*-gene product. This domain is separated from the polymerase/RNase H domain(s) located towards the carboxy-terminus by a highly variable spacer sequence (Sprengel *et al.*, 1985) which can be manipulated without loss of polymerase activity by insertions or deletions (this work and Klingmüller, 1988). Thus, as depicted in Figure 6, a multifunctional polyprotein precursor appears to be at least initially linked as such to its first DNA product. Our observation, that the connecting tether can be readily split by protease attack *in vitro* suggests that this process may also occur *in vivo* and serve to liberate the catalytic subunit(s) and to facilitate its movement along its template (Figure 6, left). However, it should be noted that a continued linkage of primer and polymerase (Figure 6, right) could also play a significant role for the primer switch from DR1 to DR2 in second strand initiation, and also in the conversion from a linear (RNA) to a circular (DNA) genome conformation. In fact, several lines of evidence argue that a TP/polymerase linkage may be present

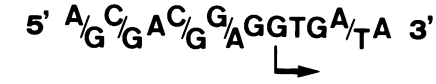

Fig. 7. Consensus sequence between direct repeat (DR) sequences DR1 from hepadna viruses. The comparison covers all sequenced hepadnaviral genomes including that of the recently discovered Heron virus (Sprengel *et al.*, 1988). It is limited to the 11 nt repeat present in the mammalian viruses. The minus-strand sequence is shown. The arrow indicates the presumed initiation site and the direction of minus-strand DNA synthesis (position 2539 in DHBV).

in at least a subfraction of the replicating viral genomes. As shown here, a significant fraction of the viral genomes can be immunoprecipitated by antisera directed against the polymerase domain (αFP6, αFP8) even after boiling in 2% SDS. Furthermore, some of these genomes have probably already completed plus-strand DNA synthesis, since they can be cleaved by restriction nucleases near the plus-strand 3'-terminus (Bosch *et al.*, 1988). In addition, the DNA polymerase activity in viral nucleocapsids was found to be tightly linked to mature viral genomes, and in particular to the DNA fragment shown to contain the genome-linked TP (Radziwill *et al.*, 1988).

The present study also provides further information on the biochemical nature of the TP–DNA linkage. Our sizing experiments (Figure 5, and further experiments, Bartenschlager, 1987) indicate that the 5'-end of the DNA minus-strand is at position 2539. Previous work by Lien *et al.* (1987) assigned the ultimate nucleotide to position 2537. However, a close examination of their experimental data (Figure 3) reveals that neither G residue 2537 nor T residue 2538 are linked to the terminal protein. Thus, the site of minus-strand initiation is at a G residue in the centre of the 12 bp repeat sequence DR1. This nucleotide is part of the GGTG sequence that is absolutely conserved between the DR sequences of all hepadnaviral genomes sequenced so far (Figure 7) which is in keeping with the notion that hepadnaviral reverse transcription initiates at this site.

With respect to the protein moiety, our data indicate that the site of DNA linkage is a tyrosine residue in the P10 segment of the TP. Furthermore, they characterize this site as being part of a small tryptic peptide. The TP primary sequence as deduced from the DHBV DNA sequence contains several tyrosine residues that meet these predictions. These candidates for a DNA linkage are presently being tested further using biochemical and genetic techniques.

Finally, it should be pointed out that the covalent coupling of the reverse transcription apparatus to a sequence-specific primer protein provides a means to target the hepadnaviral replication enzymes to their RNA genome during core assembly, and this arrangement probably also helps to prevent promiscuous reverse transcription of cellular RNAs. In retroviral replication, these aims are achieved by synthesizing the *pol*-gene product fused to the *gag* protein which self-assembles and encapsidates the retroviral genomic RNA. This is not the case in hepadnaviral replication as has been concluded from detailed investigations in the DHBV system demonstrating that the *P*-gene translation is initiated directly, and uncoupled from the overlapping core gene (Schlicht *et al.*, 1989). Thus, by fusing a protein primer to their reverse transcriptase, hepadna viruses have developed a mechanism to ensure the formation of functional nucleocapsids that is fundamentally different from that utilized by retroviruses.

Materials and methods

Viral strains and DHBV-DNA containing plasmids

Pekin ducks, predominantly infected with the European strain of DHBV (DHBV 3, Sprengel *et al.*, 1985) were used as a source for DHBV cores. All plasmids used in this report contained inserts of DHBV 16 DNA, the American strain (Mandart *et al.*, 1984). Mutational analysis was carried out with a hybrid genome of these two strains, DHBV 316, in which a *Bgl*II–*Kpn*I fragment (nt 391–1290) from DHBV 16 was substituted by the corresponding fragment of DHBV 3 (Klingmüller, 1988).

Antisera

The preparation of antisera raised against fusion proteins (FP) expressed in *Escherichia coli* or a synthetic peptide (P4), is described in detail elsewhere (Schlicht *et al.*, 1987a; Bosch *et al.*, 1988). The nucleotide positions of the coding DNA-fragment (NT), and the amino acid positions in the respective 785 amino acid long *P*-gene product (in the case of FPpreS the *pre*S-gene product) were as follows: FP6, nt 1658–2033, amino acid 496–621; FP8, nt 1087–1334, amino acids 306–388; FP10, nt 391–718, amino acids 73–183; FP11, nt 14–320, amino acids (–52) –50; peptide P4, nt 377–427, amino acids 69–86; FPpreS, nt 688–1097, amino acids (–39) –99. A core-specific antiserum (αC), raised against the full length core protein expressed unfused in *E.coli* and recognizing both DHBc- and DHBe-specific proteins, was also used (Schlicht *et al.*, 1987b).

Purification of replicative complexes (cores) from infected liver

(A) *Purification by CsCl-gradient.* Large scale preparations of core particles for analysis of viral proteins were made as follows. DHBV-infected ducks were sacrificed at 1–3 months after hatching and livers were frozen immediately and stored at –70°C. For preparation, 1 part of liver was homogenized in 2.5 parts of TPT [10 mM Tris, pH 8.0, 100 mM NaCl, 1 mM EDTA, 1 mM phenylmethylsulphonylfluoride (PMSF) and 200 KIE/ml Trasylol (Bayer, Leverkusen, FRG)] with an Ultra-Turrax at 4°C. To monitor the position of viral cores, during purification, an aliquot of already purified nucleocapsids was subjected to an endogenous polymerase reaction (see below) in the presence of radioactive deoxyribonucleoside triphosphates (dNTPs) and added to the homogenate. These cores were used as an internal marker, allowing a rapid detection of core-containing fractions. After centrifugation at 35 000 *g* for 60 min at 4°C, viral particles were concentrated by sedimentation into a step gradient, consisting of 7 ml CsCl, *d* 1.5, 7 ml CsCl, *d* 1.3, and 5 ml 30% sucrose (all in TPT) in a 45Ti rotor using 100 ml Quick-seal tubes (Beckman). Centrifugation was at 186 000 *g* for 4 h at 10°C. Fractions (1 ml) were collected from the bottom of the tube, and positive fractions containing the ^{32}P-labelled 'marker-particles' were pooled and dialysed overnight against TNE (10 mM Tris, pH 8.0, 100 mM NaCl, 1 mM EDTA). Floculated proteins were removed by centrifugation for 5 min at 6000 *g* and the supernatant was subjected to a final centrifugation into a step-gradient, consisting of 2.5 ml CsCl, *d* 1.5, 2.5 ml CsCl, *d* 1.3 and 1 ml 30% sucrose (all in TPT) in a SW40Ti rotor. After 16 h centrifugation at 217 000 *g* at 4°C, fractions were monitored as described above and shown in Figure 1 and stored in CsCl at 4°C.

(B) *Purification by sucrose gradients.* To obtain core particles with a high enzymatic activity in the endogenous polymerase reaction (Figure 5), the above method was modified by replacing CsCl-gradients by sucrose gradients. Here, the precleared crude liver extract was layered onto a sucrose step gradient, consisting of each 6 ml of 60, 30 and 15% sucrose (all in TPT). After 4 h centrifugation at 4°C at 112 000 *g* in a SW28 rotor, 500 μl fractions were collected from the bottom of the tube and the position of viral nucleocapsids monitored by dot blot analysis. Positive fractions were pooled, layered onto a continuous sucrose gradient from 60 to 15% and centrifuged at 4°C for 16 h at 217 000 *g*. DHBV DNA containing fractions (500 μl) were monitored in the same way and stored in sucrose at 4°C. Core particles prepared by this method, had a 2- to 5-fold higher activity in the endogenous polymerase reaction compared to 'CsCl-cores'.

Immunoprecipitation, protease digestion and dot blot analysis of the DNA–protein complex

Aliquots of the DNA/protein complex (fractions 1 and 2 in Figure 1) were dialysed against TE (10 mM Tris, pH 8.0, 1 mM EDTA), denatured by boiling in 1–2% SDS, 0.5–1% 2-mercaptoethanol (2-ME), diluted to RIPA (0.1% SDS, 1% Nonidet P-40 (NP-40), 0.5% deoxycholate in phosphate-buffered saline (PBS)) and subjected to an immunoprecipitation with 10 μl antiserum and 25 μl packed protein A–Sepharose (Pharmacia, Uppsala, Sweden). Immune complexes were washed four times with RIPA and once with PBS to remove detergent, and solubilized by boiling in 150 μl 2% 2-ME. Supernatants were applied to nitrocellulose filters with a dot blot apparatus and incubated under standard hybridization conditions (Maniatis *et al.*, 1982) with a ^{32}P-labelled, plus-strand DHBV DNA probe (see below). Competitions were performed by incubating the antiserum with 10 μg fusion protein or peptide at 4°C for at least 6 h, followed by the addition of the complex.

In the experiment described in Figure 3, the denatured complex was incubated with protease V8 from *S.aureus* in PBS at 37°C with an enzyme/substrate ratio of 2% (w/w). This was achieved by the addition of bovine serum albumin as carrier. Digestion was terminated by the addition of PMSF to 10 mM.

DNA probes and size markers

Plus- and minus-strand DHBV DNA probes were labelled with [α-^{32}P]-dATP (3000 Ci/mmol, Amersham-Buchler, Braunschweig, FRG) following the method of Zoller and Smith (1982). Briefly, oligonucleotides were hybridized in 20 mM Tris, pH 7.5, 10 mM $MgCl_2$, 50 mM NaCl and 1 mM DTE to DHBV plus- or minus-strand DNA, cloned into M13 vectors (Büscher *et al.*, 1985) and elongated with the Klenow enzyme in the same buffer in the presence of a single radioactive dNTP.

For preparation of the size marker (Figure 5), three colinear DHBV minus-strand oligonucleotides, starting at positions 2542, 2540 and 2538 and sharing a common 3′-end at position 2523, were synthesized on an Applied Biosystems 380A synthesizer using the phosphoramidite standard procedure by collecting aliquots of the solid phase materials after the 16th, 18th and 20th cycle. After purification on 20% polyacrylamide gels, a fraction of each oligonucleotide was phosphorylated at the 5′-end, using standard procedures (Maniatis *et al.*, 1982). After hybridization to DHBV plus-strand, phosphorylated and unphosphorylated oligonucleotides were elongated as described above, and DNA was cleaved with *Nco*I and *Taq*I according to the manufacturer's instructions.

Endogenous polymerase reaction

This method, originally described by Kaplan *et al.* (1973) for HBV, uses the enzymatic activity of the encapsidated polymerase to label viral DNA by the incorporation of radioactive dNTPs into the incomplete viral genome. DHBV core particles or DNA–protein complex were incubated in 50–100 μl 50 mM Tris, pH 8.0, 40 mM $MgCl_2$, 50 mM NH_4Cl, 1% NP-40 and 0.3% 2-ME supplemented with 12.5 μM dGTP, dCTP, TTP and 10 μCi [α-^{32}P]dATP for 30–120 min at 37°C. After addition of unlabelled dATP to 12.5 μM and continued incubation for a further 30–60 min, SDS and 2-ME were added to 2%. Viral DNA was isolated after a further incubation for 20 min at 60°C by gel filtration on a G100 column. In the experiment presented in Figure 1D, the sensitivity of the core particles or the complex to digestion with micrococcal nuclease (Boehringer Mannheim, FRG) was

tested by the addition of $CaCl_2$ to 5 mM and 5 U/ml of enzyme and 15 min incubation at 37°C followed by treatment with detergent and gel filtration. After ethanol precipitation, nucleic acids were analysed on 1.2% agarose gels, containing 0.1% SDS (Challberg *et al.*, 1980).

Preparation of ^{32}P-labelled viral DNA minus-strand fragments, carrying peptides of the terminal protein at their 5′-ends

Viral DNA contained in cytoplasmic cores (Materials and methods 'B'; Summers and Mason, 1982) was ^{32}P-labelled in the endogenous polymerase reaction and isolated by gel filtration after disruption of the nucleocapsids by SDS/2-ME treatment (see above). Aliquots of the void volume were used directly for incubation with different proteases using conditions described elsewhere (Wilkinson, 1986). Briefly, tryptic digestions were carried out with an enzyme/substrate ratio of 2% (w/w) (see above) for 1–2 h at 37°C in 100 mM Tris, pH 8.0. Digestions with proteinase K and pronase P were performed in 0.5% SDS in TE for 1 h at 37°C with an enzyme concentration of 500 μg/ml. Protease action was terminated by phenol extractions and the DNA recovered by ethanol precipitation. The labelled DNA minus-strands were annealed to plus-stranded DHBV DNAs, cloned in M13 vector (10- to 20-fold molar excess) in 20 mM Tris, pH 7.5, 1 mM EDTA and 200 mM NaCl by heating up to 100°C, followed by 2 h incubation at 60°C. Restriction endonuclease digestions were performed after the addition of an equal volume of the appropriate buffer. After phenol extraction, nucleic acids were ethanol precipitated and separated on denaturing polyacrylamide gels.

Analysis of DHBV-constructs in cell culture

DHBV constructs for cell culture were based on a pBR322-based vector, carrying a DHBV genome of 1.4 genome length (Schlicht *et al.*, 1989). All methods used to analyse gene functions have been described in detail elsewhere (Schlicht *et al.*, 1987b; Galle *et al.*, 1988). Briefly, cells were lysed 3 days after transfection in TNE/NP-40 (10 mM Tris, pH 8.0, 100 mM NaCl, 1 mM EDTA, 1% NP-40) and cytoplasmic cores isolated from the crude extract by immunoprecipitation with the core-specific antiserum. Nucleocapsids bound to the protein A–Sepharose were washed three times with lysis buffer, once with PBS and subjected to an endogenous polymerase reaction with one labelled dNTP. After digestion with proteinase K, labelled DNA was analysed on SDS-containing agarose gels.

Acknowledgements

We thank U.Klingmüller for providing DHBV 316-3δ3, M.Nassal for synthetic oligonucleotides, H.J.Schlicht for stimulating discussions, C.Kuhn for help in preparing the figures, N.Schek for a critical reading and K.Coutinho for preparing the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) SFB 229, and the Bundesministerium für Forschung und Technologie (BMFT) BCT 0381/5.

References

Bartenschlager,R. (1987) Diploma thesis, University of Heidelberg.
Bosch,V., Bartenschlager,R., Radziwill,G. and Schaller,H. (1988) *Virology*, **166**, 475–485.
Büscher,M., Reiser,W., Will,H. and Schaller,H. (1985) *Cell*, **40**, 717–724.
Challberg,M.D., Desiderio,S.V. and Kelly,T.J.,Jr (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5105–5109.
Chow,M, Bodnar,J.W., Polvino-Bodnar,M. and Ward,D.C. (1986) *J. Virol.*, **57**, 1094–1104.
Cotmore,S.F. and Tattersall,P. (1988) *J. Virol.*, **62**, 851–860.
Daubert,S.D. and Breuning,G. (1984) *Methods Virol.*, **8**, 347–379.
Galle,P., Schlicht,H.J., Fischer,M. and Schaller,H. (1988) *J. Virol.*, **62**, 1736–1740.
Ganem,D. and Varmus,H.E. (1987) *Annu. Rev. Biochem.*, **56**, 651–693.
Gerlich,W. and Robinson,W.S. (1980) *Cell*, **21**, 801–809.
Kaplan,P.M., Greenman,R.L., Gerin,J.L., Purcell,R.H. and Robinson,W.S. (1973) *J. Virol.*, **12**, 995–1005.
Klingmüller,U. (1988) Diploma thesis, University of Heidelberg.
Lien,J.M., Petcu,D.J., Aldrich,C.E. and Mason,W.S. (1987) *J. Virol.*, **61**, 3832–3840.
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*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Molnar-Kimber,K.L., Summers,J., Taylor,J.M. and Mason,W.S. (1983) *J. Virol.*, **45**, 165–172.
Radziwill,G. (1988) Ph.D. thesis, University of Heidelberg.
Radziwill,G., Zentgraf,H., Schaller,H. and Bosch,V. (1988) *Virology*, **163**, 123–132.
Rothberg,P.G., Harris,J.R.T., Nomoto,A. and Wimmer,E. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4868–4872.
Schlicht,H.J., Kuhn,C., Guhr,B., Mattagliano,R.J. and Schaller,H. (1987a) *J. Virol.*, **61**, 2280–2285.
Schlicht,H.J., Salfeld,J. and Schaller,H. (1987b) *J. Virol.*, **61**, 3701–3709.
Schlicht,H.J., Radziwill,G. and Schaller,H. (1989) *Cell*, in press.
Spandau,D. and Lee,C. (1988) *J. Virol.*, **62**, 427–434.
Sprengel,R., Kuhn,C., Will,H. and Schaller,H. (1985) *J. Med. Virol.*, **15**, 323–333.
Sprengel,R., Kaleta,E.F. and Will,H. (1988) *J. Virol.*, **62**, 3832–3839.
Summers,J. and Mason,W. (1982) *Cell*, **29**, 403–415.
Toh,H., Hayashida,H. and Miyata,T. (1983) *Nature*, **305**, 827–829.
Wilkinson,J.M. (1986) In Darbe,A. (ed.), *Practical Protein Chemistry*. John Wiley & Sons, Ltd, New York, pp. 121–148.
Wimmer,E. (1982) *Cell*, **28**, 199–201.
Zahm,P., Hofschneider,P.H. and Koshy,R. (1988) *Oncogene*, **3**, 169–177.
Zoller,M.J. and Smith,M. (1982) *Nucleic Acids Res.*, **10**, 6487–6500.

Received on September 5, 1988; revised on October 13, 1988