Localization on the viral genome and nucleotide sequence of the gene coding for the two major polypeptides of the Hepatitis B surface antigen (HBs Ag)

Patrick Charnay*, Elisabeth Mandart*, Annie Hampe*, Françoise Fitoussi*, Pierre Tiollais** and Francis Galibert*†

*Lab. Hematol. Exp., Cent. Hayem, Hop. Saint-Louis, 2 place du Dr. Fournier, 75010 Paris,
**Recombinaison et Expression Genet. (Inst. Nat. de la Santé et de la Rech. Méd. U. 163), Unité de Génie Génét., Institut Pasteur 28, Rue du Dr. Roux, 75015, Paris, France

Received 9 July 1979

ABSTRACT

The structural gene coding for both polypeptides I and II which are the two major polypeptides of the Hepatitis B surface antigen, is found to be localized on the viral genome. This gene, referred to as gene S, is located in the partially single stranded region. It maps between positions 73.6 and 95.1 % of the genome length. It is composed of 678 nucleotides, which correspond to a theoretical polypeptide of 25,422 molecular weight.

INTRODUCTION

Hepatitis B is a frequent disease mainly observed in some areas of tropical Africa and South-East Asia (1). The etiological agent is a virus (HBV), composed of an envelope and a capsid to which a DNA polymerase is associated. Its genome is a circular DNA molecule of about 3,200 nucleotides long, containing a single stranded region (2-7).

The viral envelope carries a surface antigen (HBs Ag) composed of several antigenic determinants : the "a" determinant which is group-specific and two pairs of mutually exclusive determinants : d/y and w/r (8). The electrophoretic analysis of the envelope proteins shows the presence of two to seven polypeptides. The two major ones are called : polypeptides I and II. The molecular weight of the polypeptide I was estimated between 22,000 and 26,000; the polypeptide II is glycosylated and its molecular weight was estimated between 28,000 and 30,000 (9-12). The amino-acid compositions of the two polypeptides appear very similar. Furthermore, the sequences of their nineteen first and three last amino acids are identical, suggesting that polypeptide II could differ from polypeptide I by glycosylation only (12, 13). Up to now no clear demonstration has shown that the two polypeptides are coded by the viral genome. Concerning the other polypeptides, it is difficult to know if they are coded by the viral or the cellular genome, and even if they are really part of the viral envelope.

The absence of a cellular culture system, allowing the propagation of the virus, makes its biological study particularly difficult. However, one of the possible approaches to this problem is the analysis of the primary structure of the viral DNA, which is now possible through cloning of the HBV genome in *E.coli* (14, 15, 16).

In this article we report the precise localization of the nucleotide sequence which most likely codes for polypeptides I and II and establish the primary structure of this gene, referred to as gene S.

MATERIALS AND METHODS

Enzymes and chemicals : EcoRI, BamHI, HhaI, HincII, HaeIII, XbaI, MboI, HinfI, HpaII, XhoI restriction enzymes were from Biolabs. DNA polymerase I was from Boehringer. Bacterial alkaline phosphatase and polynucleotide Kinase were from P.L. Biochemicals. Chemicals utilized were : Dimethyl sulfate (Aldrich), Hydrazine (Eastman Kodak), Acrylamide and bis-acrylamide (2 fold crystallized grade ; Serva), Dideoxynucleoside triphosphates and deoxynucleoside triphosphates (P.L. Biochemicals). Piperidine from Merck was redistilled under vacuum.

HBV DNA preparation : The entire HBV genome (*ayw* subtype) was cloned in *E.coli* (7, 14) in the following manner : Hepatitis B virions were purified by cesium cloride gradient from the plasma of a Dane particle carrier. The viral DNA was rendered fully double-stranded using the endogenous DNA polymerase and phenol extracted (14). The λgt.WES.λB (17), used as cloning vector, was digested with EcoRI endonuclease and the two arms were separated from the EcoRIλB fragment by sucrose gradient centrifugation. 30 ng of repaired HBV DNA were mixed with 500 ng of λgt DNA and EcoRI treated (0,5 unit). After heating at 65° C for 10 min. to destroy the EcoRI endonuclease the DNA was ligated with the bacteriophage T4 DNA ligase and used to transfect $CaCl_2$ treated *E.coli* C600$r_k^-m_k^-$RecBC (18). The cloned DNA is referred to as Eco HBV DNA.

The recombinant bacteriophage was amplified by plate stocks and the DNA extracted. In order to avoid any genetic drift, all stocks were made from the same lysate obtained after the first cloning step. After digestion of the bacteriophage DNA by EcoRI restriction enzyme, the Eco BHV DNA was purified by ultracentrifugation in a sucrose gradient (19).

Preparation of 5' ^{32}P labeled DNA fragments

10 to 20 picomoles of Eco HBV DNA were fully hydrolysed by different restriction enzymes according to the conditions recommended by the manufacturer. DNA fragments were dephosphoylated by alkaline phosphatase which was then inactivated by alkaline treatment and the DNA was precipitated with ethanol (20). After redissolution in spermidine buffer, DNA was labeled at its 5'ends with γ^{32}p ATP (NEN : 3,000 Ci/mM) and polynucleotide kinase (21). The DNA restriction fragments were separated by polyacrylamide gel electrophoresis, and eluted from the gels. The two labeled ends were separated by polyacrylamide gel electrophoresis, after restriction with an other enzyme or denaturation of the DNA fragment (Maxam and Gilbert personal communication).

DNA nucleotide sequence

The primary structure of double and single stranded DNA fragments was determined by the method described by Maxam and Gilbert (21 and personal communication). The chain-terminating inhibitors method developed by Sanger et al. (22) and adapted by Maat and Smith (23), for the double stranded fragments labeled at one 5' extremity was also utilized. The chemical and enzymatic reaction products were analyzed by electrophoresis in 8, 16 or 25 % acrylamide sequencing gels, one millimeter thick.

Biohazards

Containment conditions were recommended by the French National Control Committee. Growth of recombinant bacteriophages was done under L3B1* conditions. L3 is equivalent to P3 physical containment and B1 is intermediate between the EK1 and EK2 biological safety levels.

RESULTS

In order to determine if the HBV genome codes for polypeptides I and II, all the HaeIII fragments were labeled at their 5' end and substantial parts of their primary structures were determined by the Maxam and Gilbert method (21). From this, two nucleotide sequences, able to code for the amino acid sequences known to be present at the NH_2 and COOH termini of polypeptides I and II, were located respectively in the HaeIII E and the HaeIII F DNA fragments previously localized on the restriction map of the HBV genome (7). These nucleotide sequences, assumed to be the extremities of the S gene, allow us to localize this gene on the physical map of the genome between positions 73.6 and 95.1 with respect to the EcoRI restriction site (fig.1).

The nucleotide sequence between these two positions was then established, using the hydrazine dimethyl sulfate chemical degradative method and the chain-terminating inhibitors method. Among the various chemical reactions proposed by Maxam and Gilbert, we chose (1) a partial depurination by formic acid treatment and cleavage by piperidine which give bands of equal intensities for G's and A's and (2) a hydrazine reaction followed by cleavage with piperidine which gives bands of equal intensi-

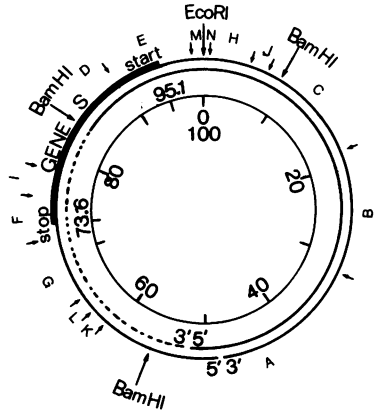

Figure 1 : Localization of gene S, coding for polypeptides I and II on the HBV genome. This diagram represents the physical map of the repaired HBV DNA (subtype ayw). The repaired region is indicated by the dotted line. The positions of Hae III restriction sites (small arrows) and BamHI restriction sites were previously determined (7). The gene S is composed of 678 nucleotides and located on the longest DNA strand. It starts in the Hae III-E fragment at position 95.1 and terminates in the Hae III-F fragment at position 73.6.

ties for C's and T's. Electrophoretic fractionation of the products of these two reactions gives, for all the bases, a band in either one or the other lane of the gel. This procedure facilitates the reading of the gel autoradiogram. To distinguish between C and T, and,A and G, the hydrazine reaction in the presence of NaCl specific for C's and the dimethyl sulfate reaction followed by piperidine cleavage, specific for G's were utilized. Moreover, a fifth reaction which gives strong bands for A's and weak bands for C's served as a control (fig. 2). In order to ensure the highest possible degree of accuracy, nucleotide sequence was determined from several overlapping fragments obtained after hydrolysis of Eco HBV DNA by BamHI, HinfI, HpaII, HaeIII and HincII restriction enzymes (fig. 3). Doing this : 1) the sequence around each restriction site used as starting points was also obtained from a fragment overlapping this site ; 2) the nucleotide sequence of each DNA strand was independently obtained. The complete nucleotide sequence of the S gene, 678 nucleotides long, is shown in figure 4.

DISCUSSION

Recently, Peterson et al. have reported that the amino acid sequences of the amino and carboxy terminal extremities of the two major polypeptides (polypeptides I and II) of the hepatitis B surface antigen are identical (12). From nucleotide sequence analysis, we are able to localize on the viral DNA, two sequences of 57 and 9 base pairs which might code for the first nineteen and last three amido acids of the polypeptides. The presence of these sequences on the viral DNA demonstrates that the gene coding for polypeptides I and II is indeed on the viral genome. The primary structure of this gene, called gene S, was then completely established. This gene, 678 nucleotides long, maps between positions 73.6 and 95.1 on the genome and is located on the longest strand in the partially single stranded region.

The reading frame corresponding to the ATG initiation codon is open until a TAA stop codon located 227 triplets downstream. The three codons corresponding to the three amino acids of the carboxy terminal extremity are in the same reading frame as the

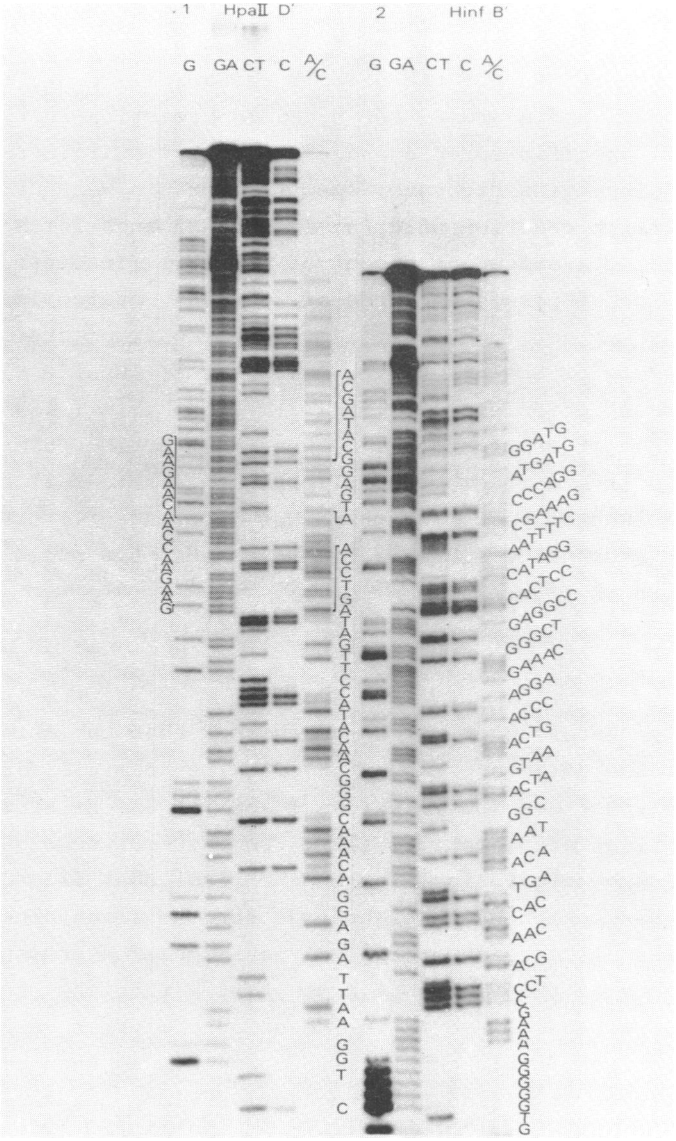

Figure 2 : Autoradiograms of sequencing gel. Products obtained by the chemical degradation method of Maxam and Gilbert (19) were fractionated by polyacrylamide gel electrophoresis. Acrylamide concentrations are 16 % for gel I and 8 % for gel 2. The positions of analysed fragments are indicated on figure 3. Gel I covers the sequence from nucleotide 334 to nucleotide 263 and gel 2 from nucleotide 566 to nucleotide 453.

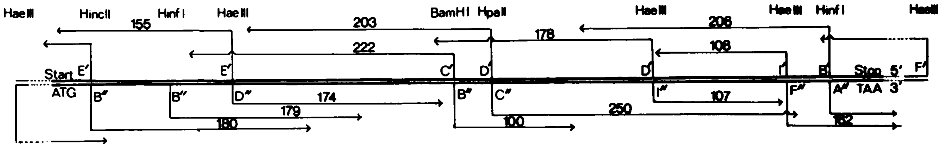

Figure 3 : Diagram of analysed DNA fragments. Vertical bars correspond to the positions of the 5' labeled ends of restriction fragments used to determine the nucleotide sequence of gene S. The numbers above each arrow indicate the length of nucleotide sequence analysed from the restriction site. The position of the restriction fragments corresponds to the physical map previously established (7). The indications "prime" and "second" correspond to the position of the labeled 5' end on the longest and the shortest viral DNA strand respectively.

ATG codon and are located immediately before this TAA codon. The second reading frame, starting from the TGG triplet contains ten stop codons (5 TAG, 4 TGA and 1 TAA) whereas the third one is open. On the other DNA strand, the three possible reading frames are closed by respectively 11, 11 and 6 stop codons. All these stop codons are distributed all along the DNA sequence.

The complete translation of the DNA starting from the ATG initiation codon would give rise to a polypeptide of 226 amino acids (fig. 4) with a molecular weight of 25,422. This figure falls within the limits of the experimental values deduced from the electrophoretic mobility in SDS polyacrylamide gels of polypeptide I and ranging from 22,000 to 26,000.

The amino acid composition deduced from the nucleotide sequence of the S gene is in reasonable agreement with the experimental data given for polypeptide I (table 1). Analysis of codon utilization does not show a preferential choice of any one.

The amino acid sequences of the amino and carboxy terminal extremities deduced from the nucleotide sequence are in perfect agreement with the experimental results given by Peterson et al. except for one amino acid (13). Peterson et al. observed in the fifteenth position the presence of a serine whereas we found a leucine. It is probably worth noting that one single base modification is sufficient to explain this difference.

The conjectured polypeptide contains a large number of proline residues (10 %) spread all along the molecule, preventing

```
                                       30                                     60
3'TAC CTC TTG TAG TGT AGT CCT AAG GAT CCT GGG GAA GAG CAC AAT GTC CGC CCC AAA AAG
5'ATG GAG AAC ATC ACA TCA GGA TTC CTA GGA CCC CTT CTC GTG TTA CAG GCG GGG TTT TTC

  met glu asn ile thr ser gly phe leu gly pro leu leu val leu gln ala gly phe phe
                                       90                                     120
  AAC AAC TGT TCT TAG GAG TGT TAT GGC GTC TCA GAT CTG AGC ACC ACC TGA AGA GAG TTA
  TTG TTG ACA AGA ATC CTC ACA ATA CCG CAG AGT CTA GAC TCG TGG TGG ACT TCT CTC AAT

  leu leu thr arg ile leu thr ile pro gln ser leu asp ser trp trp thr ser leu asn
                                      150                                     180
  AAA GAT CCC CCT TGA TGG CAC ACA GAA CCG GTT TTA AGC GTC AGG GGT TGG AGG TTA GTG
  TTT CTA GGG GGA ACT ACC GTG TGT CTT GGC CAA AAT TCG CAG TCC CCA ACC TCC AAT CAC

  phe leu gly gly thr thr val cys leu gly gln asn ser gln ser pro thr ser asn his
                                      210                                     240
  AGT GGT TGG AGA ACA GGA GGT TGA ACA GGA CCA ATA GCG ACC TAC ACA GAC GCC GCA AAA
  TCA CCA ACC TCT TGT CCT CCA ACT TGT CCT GGT TAT CGC TGG ATG TGT CTG CGG CGT TTT

  ser pro thr ser cys pro pro thr cys pro gly tyr arg trp met cys leu arg arg phe
                                      270                                     300
  TAG TAG AAG GAG AAG TAG GAC GAC GAT ACG GAG TAG AAG AAC AAC CAA GAA GAC CTG ATA
  ATC ATC TTC CTC TTC ATC CTG CTG CTA TGC CTC ATC TTC TTG TTG GTT CTT CTG GAC TAT

  ile ile phe leu phe ile leu leu leu cys leu ile phe leu leu val leu leu asp tyr
                                      330                                     360
  GTT CCA TAC AAC GGG CAA ACA GGA GAT TAA GGT CCT AGG AGT TGT TGG TCG TGC CCT GGT
  CAA GGT ATG TTG CCC GTT TGT CCT CTA ATT CCA GGA TCC TCA ACA ACC AGC ACG GGA CCA

  gln gly met leu pro val cys pro leu ile pro gly ser ser thr thr ser thr gly pro
                                      390                                     420
  ACG GCC TGG ACG TAC TGA TGA CGA GTT CCT TGG AGA TAC ATA GGG AGG ACA ACG ACA TGG
  TGC CGG ACC TGC ATG ACT ACT GCT CAA GGA ACC TCT ATG TAT CCC TCC TGT TGC TGT ACC

  cys arg thr cys met thr thr ala gln gly thr ser met tyr pro ser cys cys cys thr
                                      450                                     480
  TTT GGA AGC CTG CCT TTA ACG TGG ACA TAA GGG TAG GGT AGT AGG ACC CGA AAG CCT TTT
  AAA CCT TCG GAC GGA AAT TGC ACC TGT ATT CCC ATC CCA TCA TCC TGG GCT TTC GGA AAA

  lys pro ser asp gly asn cys thr cys ile pro ile pro ser ser trp ala phe gly lys
                                      510                                     540
  AAG GAT ACC CTC ACC CGG AGT CGG GCA AAG AGG ACC GAG TCA AAT GAT CAC GGT AAA CAA
  TTC CTA TGG GAG TGG GCC TCA GCC CGT TTC TCC TGG CTC AGT TTA CTA GTG CCA TTT GTT

  phe leu trp glu trp ala ser ala arg phe ser trp leu ser leu leu val pro phe val
                                      570                                     600
  GTC ACC AAG CAT CCC GAA AGG GGG TGA CAA ACC GAA AGT CAA TAT ACC TAC TAC ACC ATA
  CAG TGG TTC GTA GGG CTT TCC CCC ACT GTT TGG CTT TCA GTT ATA TGG ATG ATG TGG TAT

  gln trp phe val gly leu ser pro thr val trp leu ser val ile trp met met trp tyr
                                      630                                     660
  ACC CCC GGT TCA GAC ATG TCG TAG AAC TCA GGG AAA AAT GGC GAC AAT GGT TAA AAG AAA
  TGG GGG CCA AGT CTG TAC AGC ATC TTG AGT CCC TTT TTA CCG CTG TTA CCA ATT TTC TTT

  trp gly pro ser leu tyr ser ile leu ser pro phe leu pro leu leu pro ile phe phe

  ACA GAA ACC CAT ATG TAA ATT 5'
  TGT CTT TGG GTA TAC ATT TAA 3'

  cys leu trp val tyr ile stop
```

Fiqure 4

Fiqure 4 : Nucleotide sequence of HBV S gene coding for HBs Ag polypeptides I and II. The amino acid sequence deduced from the reading frame starting from the ATG initiation codon is shown. This reading frame is open until a TAA stop codon located 227 codons downstream. An other reading frame, starting with the first GGA triplet of the same DNA strand, is also open.

the existence of long alphahelical structures. The percentage of cysteine residues is also very high; they are found essentially in the central region of the molecule. 5 cys residues are loca-

Amino acid	mole %					Amino acid residues per molecule
	HBs Ag	HBs Polypeptide I				
	a(ref. 8)	b(ref.12)	c(ref.11)	d(ref.11)	e(this work)	f(this work)
Aspartic acid	5.3	5.6	7.2	5.6	1.3	3
Asparagine					2.2	5
Threonine	7.8	8.4	8.1	9.6	8.4	19
Serine	8.3	11.9	13.0	12.5	11.0	25
Glutamic acid	5.6	5.8	8.5	5.0	0.	0
Glutamine					4.0	9
Proline	11 6	12.5	10.7	11.7	9.3	21
Glycine	7.7	6.1	8.8	9.9	6.6	15
Alanine	4.0	3.5	4.6	4.0	2.2	5
Cysteine	4.8	7.2	5.7	5.8	5.8	13
Valine	5.9	4.1	4.4	4.7	4.4	10
Methionine	1.4	3.0	2.0	2.9	3.1	7
Isoleucine	6.2	5.7	4.6	5.2	6.2	14
Leucine	16.7	11.2	12.2	12.0	15.5	35
Tyrosine	1.1	2.4	0.9	1.1	2.7	6
Phenylalanine	8.0	4.9	4.3	5.2	7.1	16
Lysine	1.7	2.2	2.4	1.9	0.9	2
Histidine	0.6	0.7	0.5	0.8	0.4	1
Arginine	3.1	2.2	2.4	2.2	2.7	6
Tryptophane					5.8	13
					Total :	226
					M.W. :	25,422

Table I : The amino acid composition of the polypeptide deduced from the nucleotide sequence of the S gene is expressed in mole % (e) and in absolute number (f). These results are compared with the amino acid composition of HBs Ag (a), and HBs polypeptide I (b,c,d) obtained by several authors. Note that the percentage given in columns a,b,c and d does not take into account the percentage of tryptophane.

ted between positions 137 and 149. Several hydrophobic regions are observed, the largest one being located between residues 80 and 98. The nucleotide sequence, which allows the distinction between the acidic amino acids and their amid derivatives, indicates the presence of 0 glutamic acid and 3 aspartic acid residues (table I). These residues and the ten basic amino acid residues are scattered all along the molecule. On the other hand, the knowledge of the amino acid sequence of the polypeptide allows the chemical synthesis of fragments of the molecule to study their antigenic determinants.

It has been proposed that polypeptide II derives from polypeptide I by glycosylation (13). The difference in the molecular weights of these two polypeptides suggests the existence of a large amount of glycosydic residues on polypeptide II (24). This can be compared with the large amount of alcoolic amino acids : serine (11 %) and threonine (8.4 %) residues. Moreover, while there are five asparagine residues in the molecule, three of them belong to the tripeptide Asn-X-ser (or thr), which has been reported to be necessary for the formation of the N-glycosydic bound (25).

Although the molecular weight calculated from the electrophoretic mobility of polypeptide I (9-12) and the molecular weight deduced from the DNA sequence coding for that polypeptide are in good agreement, one cannot eliminate the existence of a small intervening sequence (26-27) within gene S. Nevertheless, in that case, the intervening sequence could not exceed 100 nucleotides. The comparaison of the amino acid composition of polypeptide I as determined by several authors (table I) with the amino acid composition deduced from the nucleotide sequence does not give further information concerning the existence of this eventual intervening sequence. On the other hand, glycosylation is not necessarily the only difference between polypeptide I and II. One could imagine that in the maturation process a fraction of the mRNA molecules undergoes some splicing, thus giving rise to two different final gene products as it has already been observed for SV40 (28, 29).

As mentioned above, among the five other possible reading frames, only one, present on the same DNA strand as gene S, is open. The absence of a stop codon in such a long nucleotide sequence (678 residues) might indicate that this DNA fragment could code for another polypeptide. The existence of overlapping genes which have been shown to occur in ϕX74, G4 and SV40 genomes (28-31) are worth considering since the HBV DNA is so far the smallest known animal virus genome. More information on this point will probably be obtained from the complete nucleotide sequence of the HBV genome.

ACKNOWLEDGEMENTS

We are grateful to Dr. A. FRITSCH for helpful discussion and Dr. R. WAHLEN for reviewing the manuscript. We thank Mrs M. GOBET for her excellent technical assistance.

This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM) and the Centre National de la Recherche Scientifique (CNRS).

†To whom reprint requests should be sent.

REFERENCES

1. Blumberg, B.S. (1977) Science 197, 17-25.
2. Dane, D.S., Cameron, C.H. and Briggs, M. (1970) Lancet i 695-698.
3. WHO technical report series, number 602 (1976).
4. Summers, J., O'Connel, A. and Millman, I. (1975) Proc. Nat. Acad. Sci. USA 72, 4597-4601.
5. Hruska, J.F., Clayton, D.A., Rubenstein, J.L.R. and Robinson, W.S. (1977) J. Virol. 21, 666-682.
6. Landers, T.A., Greenberg, H.B. and Robinson, W.S. (1977) J. Virol. 23, 368-376.
7. Charnay, P., Pourcel, C., Louise, A., Fritsch, A. and Tiollais, P. (1979) Proc. Nat. Acad. Sci. USA 76, 2222-2226.
8. Dreesman, G.R., Hollinger, F.B., Surians, J.R, Fujioka, R.B., Brunschwig, J.P. and Melnick, J.L. (1972) J. Virol 10, 469-476.
9. Gerin, J.L. (1974) in mecanisms of virus disease ed. W.S. Robinson, C.R. Fox pp 215-24 Menlo Park : W.A. Benjamin.
10. Dreesman, G.R., Chairez, R., Suarez, M., Hollinger, F.B., Courtney, R.J. and Melnick, J.L. (1975) J. of Virology 16, 508-515.
11. Shih, J.W. and Gerin, J.L. (1977) J. of Virol. 21, 1219-1222.
12. Peterson, D.L., Roberts, I.M. and Vyas, G.N. (1977) Proc. Nat. Acad. Sci. USA 74, 1530-1534.
13. Peterson, D.L., Chien, D.Y., Vyas, G.N., Nitecki, D. and Bond, H. (1978) in Viral Hepatitis, ed. G. Vyas, S. Cohen and R. Schmid, pp 569-573 The Franklin Institute Press, Philadelphia.
14. Fritsch, A., Pourcel, C;, Charnay, P. and Tiollais, P. (1978) C.R. Acad. Sci. Paris 287, 1453- 1546.
15. Burrell, C.J., Mackay, P., Greenaway, P.J., Hofschneider, P.H. and Murray, K. (1979) Nature 279, 43-47.
16. Sninsky, J.J., Siddiqui, A., Robinson, W.S. and Cohen, S.N. (1979) Nature 279, 346-348.
17. Leder, P., Tiemeier, D. and Enquist, L. (1977) Science 196, 175-177.
18. Tiollais, P., Perricaudet, M., Pettersson, U. and Philipson, L. (1976) Gene 1, 49-63.
19. Hérissé, J., Courtois, G. and Galibert, F. (1978) Gene 4 279-294.
20. Kroeker, W.D. and Laskowski, M.S.R. (1977) Anal. Biochem. 79, 63-72.
21. Maxam, A.M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. USA 74, 560-564.
22. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467.
23. Maat, J. and Smith, A.J.W. (1978) Nucleic Acid. Res. 5, 4537-4545.
24. Shiraishi, H. Kohama, T., Shirachi, R. and Ishida, N. (1977) J. Gen. Virol. 36, 207-210.
25. Struck, D.K., Lennarz, W.J. and Brew, K. (1978) J. Biol. Chem. 253, 5786-5794.
26. Berget, S.M., Moore, C. and Sharp, P.A. (1977) Proc. Nat. Acad. Sci. USA 74, 3171-3175.
27. Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, J. (1977) Cell 12, 1-8.

28. Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J. Celma, C.L. and Weissman, S.M. (1978) Science 200, 494-502.
29. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert , G. and Ysebaert, M. (1978) Nature 273, 113-117.
30. Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchinson III, C.A., Slocombe, P.M. and Smith, M. (1977) Nature 265, 687-691.
31. Barrell, B.G., Shaw, D.C., Walker, J.E., Northrop, F.D., Godson, G.N. and Fiddes, J.C. (1978) Biochem. Soc. Trans. 6, 63-67.