

**Table 1** Areas enclosed by maximum grain size isopleths ( $\times 10^3 \text{ km}^2$ )

	Maximum pumice (mm)				
	64	32	16	8	4
Rotoehu bed G*	0.2	3.4	14	24	44
Taupo plinian	5	9.9	18	28†	40†
Pompei plinian <sup>23</sup>	0.3	0.9	1.9†	3†	—
	Maximum lithic (mm)				
	32	16	8	4	
Rotoehu bed G*	0.1	2.5	12	25	
Taupo plinian	2.3	4.2	7	15	
Pompei plinian <sup>23</sup>	0.6	1.1†	1.6†	—	

\* Assuming a symmetrical distribution on either side of the Te Puke-Te Teko line.

† Extrapolated values.

uniformly in all directions. The material which fell 100 km out from source has a median terminal fall velocity of  $3 \text{ m s}^{-1}$ , for which the fall time from a height of 10–50 km in normal atmosphere is 35–70 minutes. This suggests that the outward radial velocity of the expanding cloud averaged  $180\text{--}90 \text{ km h}^{-1}$  neglecting the dustiness of the atmosphere at the time and turbulence. The lateral expansion was apparently much greater than is normal in volcanic explosions, and was comparable with the most powerful one currently known. In normal explosions,

the expansion is due to magmatic gases which constitute a few wt% of the magma. In the Rotoehu case it is possible that steam constituted a few tens of wt% of magma involved.

This ash is an example of a new class of pyroclastic deposit for which the name 'littoral co-ignimbrite ash-fall' is proposed.

I thank the Royal Society of New Zealand for supporting this research, which was done at the University of Auckland while I was on leave from Imperial College London, and my wife for assistance.

Received 3 July; accepted 21 August 1979.

1. Healy, J., Vucetich, C. G. & Pullar, W. A. *N.Z. Geol. Surv. Bull.* n.s. 73 (1964).
2. Vucetich, C. G. & Pullar, W. A. *N.Z. J. Geol. Geophys.* 16, 745–780 (1973).
3. Vucetich, C. G. & Howarth, R. *N.Z. J. Geol. Geophys.* 19, 51–69 (1976).
4. Howarth, R. *N.Z. J. Geol. Geophys.* 18, 683–712 (1975).
5. Vucetich, C. G. & Pullar, W. A. *N.Z. J. Geol. Geophys.* 12, 784–837 (1969).
6. Nairn, I. A. *N.Z. J. Geol. Geophys.* 15, 251–261 (1972).
7. Thompson, B. N. *N.Z. J. Geol. Geophys.* 11, 1189–1191 (1968).
8. Pullar, W. A. & Nairn, I. A. *N.Z. J. Geol. Geophys.* 15, 446–450 (1972).
9. Sparks, R. S. J. & Walker, G. P. L. *J. Volcan. geotherm. Res.* 2, 329–341 (1977).
10. Walker, G. P. L. *Contr. Miner. Petrol.* 36, 135–146 (1972).
11. Wright, J. V. & Walker, G. P. L. *Geology* 5, 729–732 (1977).
12. Sparks, R. S. J., Self, S. & Walker, G. P. L. *Geology* 1, 115–118 (1973).
13. Walker, G. P. L., Self, S. & Froggatt, P. C. (in the press).
14. Kuno, H. *et al. Jap. J. Geol. Geogr.* 35, 223–238 (1964).
15. Walker, G. P. L. *J. Volcan. Geotherm. Res.* (in the press).
16. Booth, B., Croasdale, R. & Walker, G. P. L. *Phil. Trans. R. Soc. A* 288, 271–319 (1978).
17. Cole, J. W. & Stephenson, T. M. *Publ. geol. Dept. Victoria Univ. Wellington*, 1 (1972).
18. Topping, W. W. *N.Z. J. Geol. Geophys.* 16, 397–423 (1973).
19. Thorarinnsson, S. *Bull. Volcan.* 16 1–44 (1953).
20. Fisher, R. V. *Geol. Rdsch.* 57, 837–864 (1968).
21. Verbeek, R. D. M. *Krakatau* (Batavia, 1885).
22. Stehn, C. E. *Proc. 4th Pacif. Sci. Congr.* 1–55 (1929).
23. Lirer, L., Pescatore, T., Booth, B. & Walker, G. P. L. *Bull. geol. Soc. Am.* 84, 759–772 (1973).

## Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in *E. coli*

Francis Galibert\*†, Elisabeth Mandart\*, Françoise Fitoussi\*, Pierre Tiollais‡ & Patrick Charnay\*§

\* Laboratoire d'Hématologie Expérimentale, Centre Hayem, Hôpital Saint-Louis, Paris, France

‡ Recombinaison et Expression Génétique (INSERM U.163), Unité de Génie Génétique, Institut Pasteur, Paris, France

*The complete nucleotide sequence of hepatitis B virus genome (subtype ayw) cloned in Escherichia coli has been determined using the Maxam and Gilbert method and the dideoxynucleotide method. This sequence is 3,182 nucleotides long. Location of the nonsense codons shows that the coding capacity of the L chain is larger than the coding capacity of the S chain. Eight open regions, able to code for polypeptide chains larger than 100 amino acids, have been located. Region 6, which is the largest, covers more than 80% of the genome. The gene S which codes for polypeptide I of the Hbs Ag and was previously located between coordinates 95.1 and 73.6 is contained in region 7.*

HEPATITIS B VIRUS (HBV) infects man at a very high rate. In western Europe and USA, about 10% of the population carry an HBV serological marker such as the surface antigen (HBs Ag) or the corresponding antibody (HBs Ac)<sup>1</sup>. In some areas of tropical Africa and southeastern Asia, this percentage can be as high as 60% or 70% (ref. 2). Recovery is spontaneous in most cases, but the virus can cause very severe complications like fulminant hepatitis, chronic hepatitis and cirrhosis in addition to which a hepatocarcinoma can develop<sup>3</sup>. It has been estimated that there

are at least 120 million chronic carriers of HBs Ag, the so-called healthy carriers, throughout the world<sup>2</sup>.

In addition to HBs Ag, two other antigens have been identified: HBc Ag which corresponds to the core of the virus, and HBe Ag whose location within the virion has been recently demonstrated<sup>4</sup>. It was also shown that the viral genome is made up of a circular DNA molecule which is partially single-stranded and about 3,200 bases long<sup>5,6</sup>. The virion also harbours a DNA polymerase able to repair the viral DNA<sup>7</sup>. The novel structure of this virus makes its classification somewhat difficult.

The biological study of the virus is complicated by the fact that HBV cannot be propagated in cell cultures and the only available viral source remains human serum. However, cloning of the HBV genome in *Escherichia coli* has allowed the purification of large amounts of genetically homogeneous HBV DNA, and the establishment of several restriction maps<sup>8–10</sup>. Hence we have recently<sup>11</sup> localised and determined the primary structure of the S gene which codes for polypeptide I and its glucosylated derivative polypeptide II, which are the major components of the HBs Ag.

In this article, we report the complete primary structure of the viral genome (subtype ayw) previously cloned in *E. coli*.

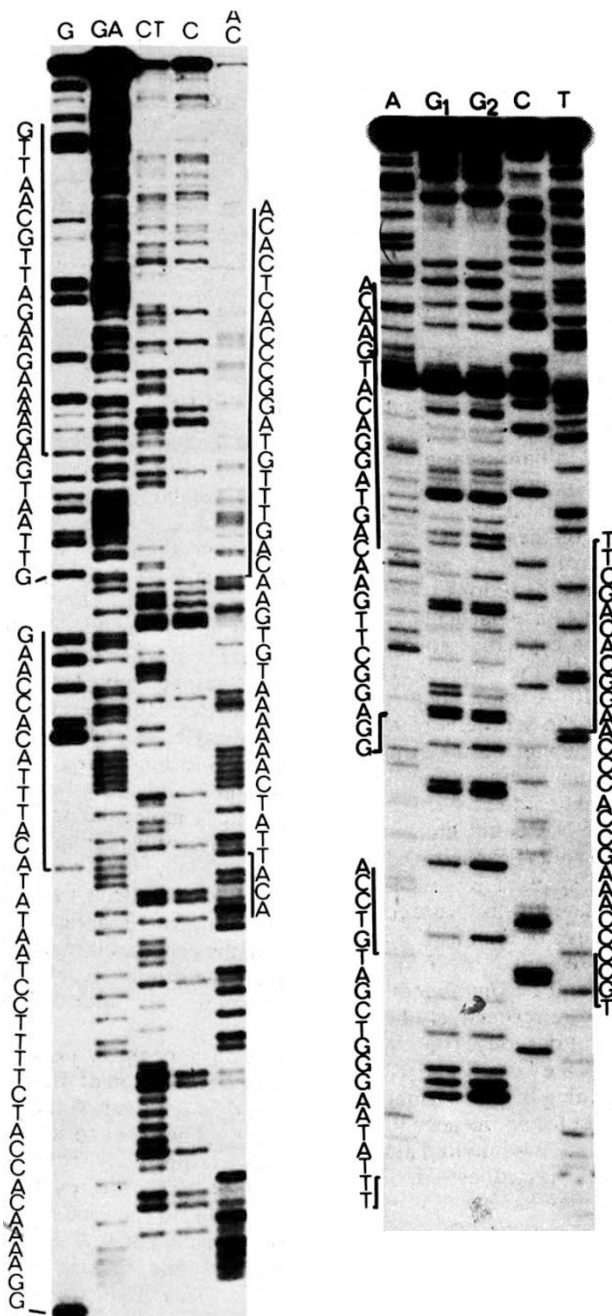
### Primary structure of the cloned viral genome

To determine the nucleotide sequence of this DNA, the chemical degradation method of Maxam and Gilbert<sup>12,13</sup> and the dideoxynucleotide method of Sanger *et al.*<sup>14</sup> as modified for

† To whom correspondence should be addressed.

§ Present address: Unité de Génie Génétique, Institut Pasteur, Paris, France.

**Fig. 1** Viral DNA (serotype ayw), repaired with the endogenous DNA polymerase and the four nucleotide triphosphates, was digested with *EcoRI* endonuclease and cloned in *E. coli* using  $\lambda$  gt WES bacteriophage as the cloning vehicle<sup>16</sup>. The cloned DNA was referred to as *Eco* HBV DNA. To avoid genetic drift, all stocks used to analyse the DNA sequence were made on plates from the same lysate obtained after the first cloning step. Recombinant phages were purified by two successive CsCl gradients and the DNA was phenol extracted<sup>11</sup>. After hydrolysis with *EcoRI* endonuclease, *Eco* HBV DNA was separated from the two  $\lambda$  arms by sucrose gradient ultracentrifugation<sup>17</sup>. To prepare each set of restriction DNA fragments, 25 to 30  $\mu$ g of *Eco* HBV DNA were fully hydrolysed with one of the restriction enzymes described in Fig. 2, using the conditions recommended by the manufacturer (Biolabs). Restriction fragments were dephosphorylated by incubation at 65 °C for 45 min with 4  $\mu$ l of bacterial phosphatase. This enzyme was then inactivated by alkaline treatment<sup>18</sup> and DNA fragments were 5' labelled with [ $\gamma$ -<sup>32</sup>P]ATP (NEN; specific activity 2,500 Ci mmol<sup>-1</sup>) and polynucleotide kinase (P. L. Biochemicals) as described by Maxam and Gilbert<sup>12</sup>. Labelled fragments were fractionated by gel electrophoresis and eluted from the gels as previously described. The two labelled ends of each restriction fragment were then separated either by hydrolysis of the DNA fragment with another restriction endonuclease and gel electrophoresis or by heat denaturation in the presence of 30% dimethyl sulfoxide and gel electrophoresis<sup>13</sup>. DNA fragments labelled at one 5' end were eluted from the preparative gel and dissolved in water. Chemical reaction, specific for G, A + G, C + T, T and A + C giving a stronger band for A than C were performed as described previously<sup>12,13,17</sup>. Sequence reaction products dissolved in loading solution were heat denatured and fractionated on 25, 16 and 8% acrylamide gels. Figure 1a is an example of fractionation on an 8% acrylamide slab gel of the sequence reaction products obtained with *Hinf* C' DNA fragment (nucleotides 658 to 552). When the two labelled ends were separated by hydrolysis with a restriction enzyme, nucleotide sequences were also determined with the dideoxynucleotide method<sup>14</sup> adapted by Maat and Smith<sup>15</sup> for double-stranded DNA fragments. Figure 1b is an example of fractionation on a 16% acrylamide gel of such reaction products obtained with *Bgl*A' fragment (nucleotide sequence from residue 1,264 to residue 1,338). The two G lanes correspond to two different reaction conditions. In G1 there is no dGTP while in G2 the final concentration of dGTP is 2  $\mu$ M. In both reactions, there is 1 mM ddGTP, final concentration. Subscript prime indicates the position of the <sup>32</sup>P on the L strand, subscript second, its position on the S strand.



double-stranded DNA by Maat and Smith<sup>15</sup> were used. Two examples of sequencing gel autoradiograms are shown in Fig. 1. To analyse independently both DNA strands and to overlap all the restriction sites used as starting points, a large number of fragments were analysed (Fig. 2). The localisation of the 5' end of the largest DNA strand on the *Bam*HI viral restriction map and the knowledge of this restriction map on the *Eco* HBV DNA allowed the identification of the cloned DNA strands with respect to the viral DNA strands. Here we call the DNA strand of *Eco* HBV DNA homologous to the longest viral DNA strand the L strand, and the other, the S strand. The sequence shown in Fig. 3 corresponds to the *Eco* HBV DNA, which is 3,182 nucleotides long.

### Location of the open reading frames on the viral genome

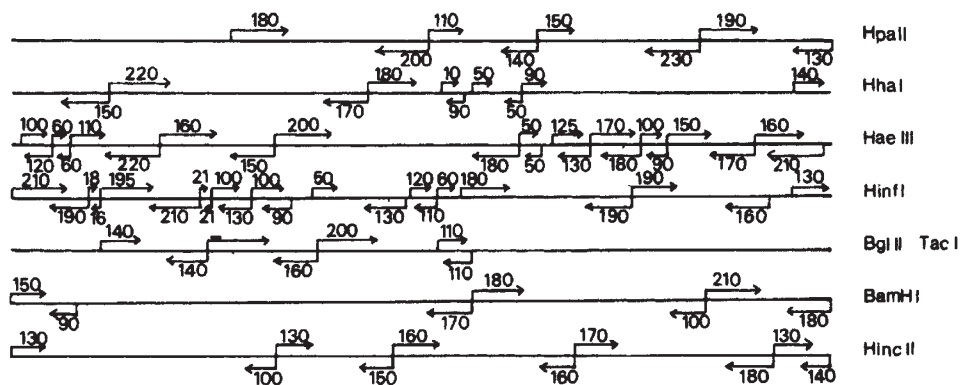
A computer program was used to search for nonsense codons on both DNA strands<sup>19</sup>. It appears from Fig. 4 that the number and distribution of nonsense codons are very different on S and L strands. Considering the total genome length, a random distribution of nucleotides should give about 50 nonsense codons in each reading frame. Instead one observes 33, 29 and 15 nonsense codons in the S strand for each phase and 61, 60 and 41 nonsense codons on the L strand. In addition, their distribution along the DNA molecule is very different on the two strands. On the L strand, the open reading frames are relatively short (Fig. 4). Four regions are able to code for 100 amino acids or more. These four regions labelled 1 to 4 are 510, 327, 447 and 345

nucleotides long. On the S strand (Fig. 4), the open regions are much longer. These regions labelled 5 to 8 are 498, 2,553, 1,296 and 657 nucleotides long. The length of the above hypothetical coding regions which include the *Eco*RI restriction site (such as regions 4, 6 and 7) is only valid if no DNA has been lost in the vicinity of this restriction site on cloning, which is most likely. In spite of this limitation, the distribution of stop codons clearly shows that the coding capacity of the L strand is larger than that of the S strand.

Regions 1, 2, 3 and 4 appear to be located on the genome as two pairs (Fig. 5)—region 3 is located immediately downstream of region 1 and region 2 immediately beyond region 4. Regions 5, 6, 7 and 8 extend over the whole genome and overlap with each other in different reading frames (Fig. 5). Region 6 covers about 80% of the genome and includes the whole of region 7 and part of regions 5 and 8. Regions 6 and 7 correspond partly to the single-stranded region of the genome. Regions 4 and 5 cover the 5' extremity of the S strand and the nick of the L strand.



**Fig. 2** Diagram of analysed DNA fragments. Vertical bars correspond to the position of the 5' labelled ends of restriction fragments used to determine the nucleotide sequence of the *Eco* HBV DNA. Numbers above each arrow indicate the length of nucleotide sequences analysed from the restriction sites.



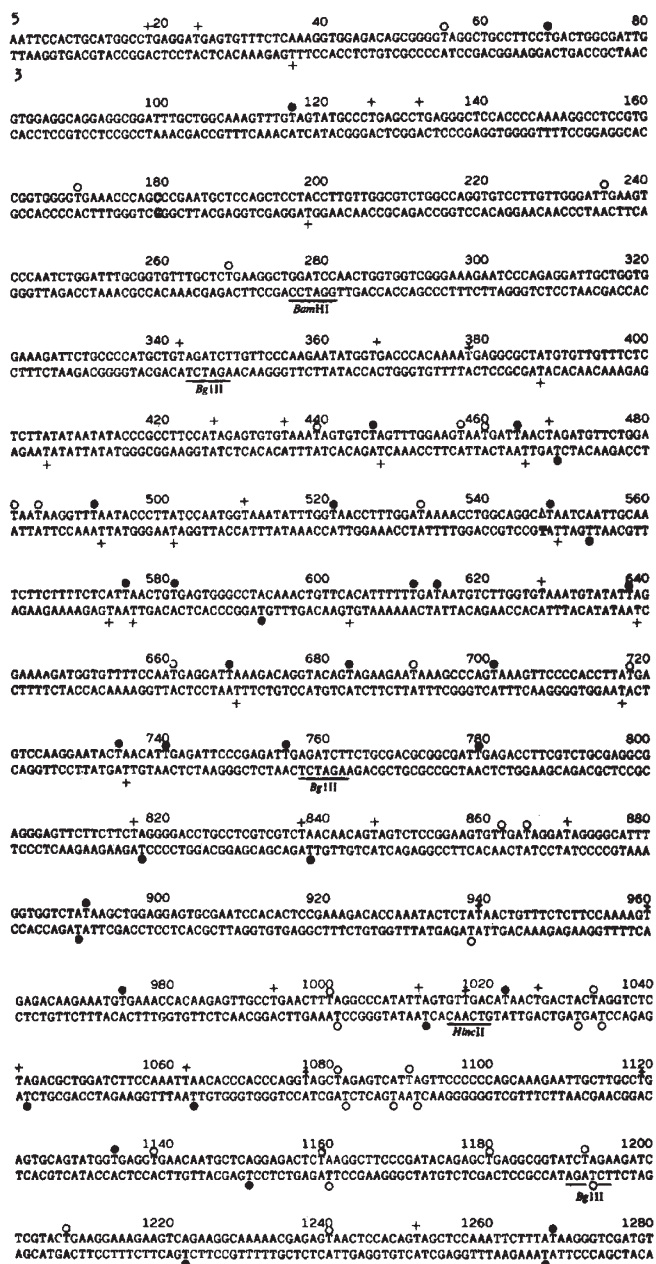
## Tentative translation of open regions

Few proteins have been shown to be coded by the viral genome<sup>9,11</sup>, but one can suppose that, because of its very small size, regions with open reading frames are largely used for translation. Therefore, a search for the ATG codons which could initiate translation has been made near the beginning of these regions, although the existence of mRNA splicing<sup>20,21</sup> could complicate the translation of the open regions. The positions of these ATGs are shown by small arrows in Fig. 5. In region 1 to 4, the numbers of codons located downstream from these ATG codons are 79, 60, 145 and 28 respectively. Regions 2 and 4 have particularly low coding capacities. However, these coding capacities could be modified by mRNA splicing. For example, the translation from the first ATG located in region 4, of mRNA transcribed from regions 4 and 2 and spliced could allow the synthesis of a polypeptide with a molecular weight (MW) up to 18,000. For simplicity and clarity, only regions with an open reading frame able to code for 100 amino acids or more are taken into account. But a more complicated splicing pattern of mRNA could link, in one open reading frame, areas with small coding capacity scattered all along the genome. If region 3 which spans in the single-stranded part of the genome is actually translated into protein, this can happen only after DNA synthesis of the missing strand.

In regions 5, 6, 7 and 8, the situation appears to be very different. The number of codons located downstream of the first ATG are respectively 154, 832, 389 and 212. Region 6 (2,553 nucleotides long) contains 9 ATG codons. The first and second are located at 57 and 1,296 nucleotides respectively from the beginning of this region. The seven other ATGs are regularly distributed in the second half of this region. In the case where translation is initiated at the first ATG, a protein of MW 90,000 might be synthesised. DNA polymerases and reverse transcriptases have a MW of that order<sup>22,33</sup>, so it is tempting to suggest that the DNA polymerase found within the virion could be coded by this region.

By correlating the amino acid sequences of both ends of the HBs Ag polypeptide I<sup>24</sup> and the nucleotide sequence of the *Eco* HBV DNA, we are able to localise the gene coding for this polypeptide in region 7. This gene, called *S* gene, was previously located between positions, 95.1 and 73.6 on the genome map<sup>11</sup>. The ATG (position 3,030) corresponding to the N-terminal extremity of the polypeptide I is the third of region 7. It is located 618 nucleotides from the beginning of this region, whereas the two other ATGs (positions 337 and 13) are at 129 and 453 nucleotides from the beginning. The fact that the ATG 3030 is preceded by two other ATGs in the same open reading frame deserves some additional comment. Three explanations can be given: (1) Region 7 is completely translated and gives rise to a long polypeptide different from polypeptide I but whose C-terminal amino acid sequence is identical to polypeptide I. A similar situation occurs with protein VP2 and VP3 of SV40<sup>25,26</sup> and the 33,000 and 100,000 MW proteins of adenovirus<sup>27</sup>. (2) Region 7 is completely translated into a precursor protein of polypeptide I. (3) The region upstream of the third ATG is not translated. At the present time, one cannot choose between the

**Fig. 3** (Below and opposite) Nucleotide sequence of *Eco* HBV DNA. Stop codons in the various reading frames are indicated in the following manner: ○ stop codon in phase 1; ● in phase 2; + in phase 3. The position of *HincII*, *BamHI*, *BglII* and *XhoI* restriction sites are indicated.



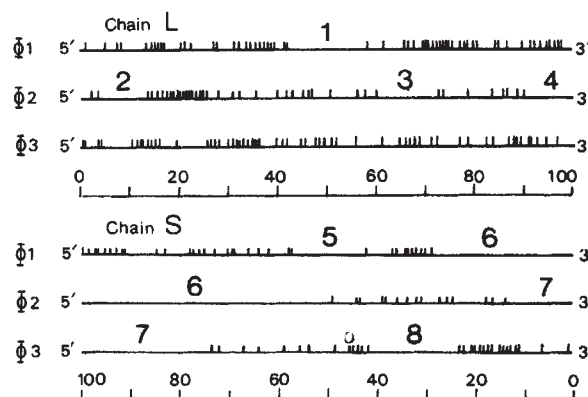
[illegible]

three hypotheses. Nevertheless, the first hypothesis could partially explain the presence of large polypeptides found in small amount in the HBs Ag preparations<sup>24</sup>.

Translation of the terminal nucleotide sequence of region 8 would give rise to a polypeptide having at its C-terminal end a repeated sequence of 9 amino acids: Ser-Pro-Arg-Arg-Arg-Arg-Ser-Glu-Ser. The last amino acid of the first sequence, Ser, would be the first one of the second sequence. In the corresponding DNA sequence, however, the only repetition is TCTCAA (Ser-Glu). Beyond the ATG codon, region 8 contains 24 arginine codons—11% of the total codons. In the last 35 codons there are 16 arginine codons. This large proportion of basic amino acids is similar to that observed in histone, which suggests that the translational product of region 8 could be a polypeptide with a high affinity for HBV DNA. Therefore, if this region is translated, one could envisage that the HBc Ag could be coded by this DNA sequence.

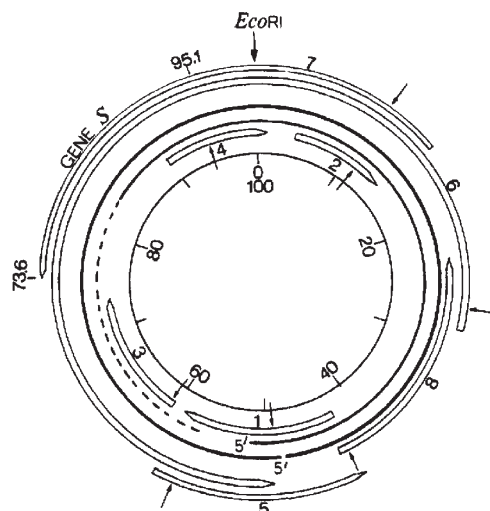
## Search for signal sequences

From a comparative analysis of the primary structure of viral or cellular mRNAs and of chromosomal DNA fragments, some remarkable DNA sequences were found. In several circumstances, a sequence identical or similar to 5'TATAAA was observed in front of the transcribed region (ref. 28 and D. Hogness and M. Goldberg, personal communication), while the sequence AAUAAA has been found at the 3' end of mRNA, near the poly(A)tail<sup>29-31</sup>. On the other hand, it has been shown that during the initiation of translation the 3'OH end of 16S ribosomal RNA hybridises to the 5' end of prokaryotic mRNA through the CCUCC sequence<sup>32</sup>. The occurrence of a similar phenomenon has not been demonstrated in eukaryotic cells. Nevertheless, sequences able to hybridise to the 3'GCGGAAGGA sequence of the 18S ribosomal RNA have been observed in some eukaryotic mRNAs<sup>28,33-35</sup>. Therefore, we looked for these sequences in the *Eco* HBV DNA. The 5'TATAAA sequence was found only at position 1,269 on the S strand at 335 nucleotides before region 6. It could therefore correspond to a signal relating to the initiation of transcription of this very large open region. In that case, since this sequence is followed by three ATGs (1,089, 1,022, 998) all in a closed reading frame, a splicing of the corresponding mRNA should



**Fig. 4** Diagram showing the localisation of the nonsense codons on chains L and S of the *EcoHBV* DNA. Three reading frames, called  $\Phi 1$ ,  $\Phi 2$  and  $\Phi 3$ , were defined from the 5' end of each DNA strand. On the L chain,  $\Phi 1$  is defined by its first triplet AAT,  $\Phi 2$  by ATT and  $\Phi 3$  by TTC. On the S chain,  $\Phi 1$  is defined by its first triplet CCA,  $\Phi 2$  by CAC and  $\Phi 3$  by ACA. The viral DNA is circular and its length in nucleotides (3,182) is not a multiple of 3. Therefore, going through the *EcoRI* site phase  $\Phi 1$  is change to phase  $\Phi 2$ , phase  $\Phi 2$  to phase  $\Phi 3$  and phase  $\Phi 3$  to phase  $\Phi 1$ . Vertical bars represent stop codons. Numbers 1 to 8 define the areas with an open reading frame arbitrarily chosen with a length above 100 codons. Region 1 goes from nucleotide 1,351 to 1,860, region 2 from nucleotide 119 to 445, region 3 from 1,904 to 2,350, region 4 from 2,855 to 17, region 5 from 1,847 to 1,350, region 6 from 935 to 1 and from 3182 to 1565, region 7 from 466 to 1 and from 3,182 to 2353, region 8 from 1,392 to 736.





**Fig. 5** Localisation of the open reading frames on the viral genome. The structure of the genome is shown as previously described<sup>8</sup>. The *EcoRI* site is the origin of the physical map. Block arrows define the open reading frames on the two viral DNA strands. The small arrows indicate the ATG codon located closest to the beginning of the open region. The *S* gene is located between positions 95.1 and 73.6 and is included in region 7.

take place to link its 5' end to a nucleotide sequence corresponding to the open reading frame of region 6. Sequences which differ by only one nucleotide were found 36 times on the S strand and 45 times on the L strand. The 5' AATAAA sequence occurs once on the L strand at position 689, 243 nucleotides after the stop codon of region 2. This sequence is not found in chain S even at the end of gene *S* which is known to code for polypeptide 1. Two sequences, 5'GCCTTCCT and CGCCTTCC, able to hybridise with the 3' end of 18S RNA ribosomal, were found at positions 61 and 417 on the L strand. Sequences able to hybridise with 6 or 7 nucleotides of the 3'OH end of this RNA were found once on the L chain and 8 times on the S chain. Among these, the CCTTCC sequence located at position 3,176 on the S strand follows a short palindromic sequence 5'GTGGAATTCCAC 3' starting at position 8. This

feature could be related to the expression of the *S* gene located downstream.

Containment conditions for this study were recommended by the French National Control Committee. Growth of recombinant bacteriophages was done in L3B1 conditions. We thank Dr A. Fritsch for helpful discussion and reviewing the manuscript, Dr R. Studen for computer work, Dr A. Maxam for sending us the revised version of the Maxam and Gilbert sequencing method before publication, and Mrs Gobet for technical assistance. This work was supported by INSERM and CNRS.

Received 7 August; accepted 14 September 1979.

1. Szmuness, W., Harley, E. S., Ikram, H. & Stevens, C. E. in *Viral Hepatitis* (eds Vyas, G. N., Cohen, S. N. & Schmid, R.) (Franklin Institute, Philadelphia, 1978).
2. Szmuness, W. *Am. J. Path.* **81**, 629-649 (1975).
3. Redeker, A. *Am. J. med. Sci.* **270**, 9-16 (1975).
4. Ohori, H., Omadera, S. & Ishida, N. *J. gen. Virol.* **43**, 423-427 (1979).
5. Summers, J., O'Connell, A. & Millman, I. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4597-4601 (1975).
6. Robinson, W. S. *A. Rev. Microbiol.* **31**, 357-377 (1977).
7. Landers, T. A., Greenberg, H. B. & Robinson, W. S. *J. Virol.* **23**, 368-376 (1977).
8. Charnay, P., Pourcel, C., Louise, A., Fritsch, A. & Tiollais, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2222-2226 (1979).
9. Burrell, C. J., Mackay, P., Greenaway, P. J., Hofschneider, P. H. & Murray, K. *Nature* **279**, 43-47 (1979).
10. Sninsky, J. J., Siddiqui, A., Robinson, W. S. & Cohen, S. N. *Nature* **279**, 346-348 (1979).
11. Charnay, P. *et al. Nucleic Acids Res.* **7**, 335-346 (1979).
12. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
13. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* (in the press).
14. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
15. Maat, J. & Smith, A. J. W. *Nucleic Acids Res.* **5**, 4537-4545 (1978).
16. Fritsch, A., Pourcel, C., Charnay, P. & Tiollais, P. *C. r. hebdom. Séanc. Acad. Sci., Paris* **287**, 1453-1456 (1978).
17. Galibert, F., Hérisse, J. & Courtois, G. *Gene* **6**, 1-22 (1979).
18. Kroecker, W. D. & Laskowski, M. S. R. *Analyt. Biochem.* **79**, 63-72 (1977).
19. Staden R. *Nucleic Acids Res.* **4**, 4037-4051 (1977).
20. Berget, S. M., Berk, A. J., Harrison, T. & Sharp, P. A. *Cold Spring Harb. Symp. quant. Biol.* **42**, 523-530 (1977).
21. Broker, T. R. *et al. Cold Spring Harb. Symp. quant. Biol.* **42**, 531-553 (1977).
22. Richardson, C. C. A. *Rev. Biochem.* **38**, 795-840 (1969).
23. Wu, A. M. & Gallo, R. C. *CRC Crit. Rev. Biochem.* **3**, 289-347 (1975).
24. Peterson, D. L., Roberts, I. M. & Vyas, G. N. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1530-1534 (1977).
25. Fiers, W. *et al. Nature* **273**, 113-120 (1978).
26. Reddy, V. B. *et al. Sciences* **200**, 494-402 (1978).
27. Axelrod, N. *Virology* **87**, 366-383 (1978).
28. Baker, C. C., Hérisse, J., Courtois, G., Galibert, F. & Ziff, E. *Cell* (in the press).
29. McReynolds, L. *et al. Nature* **273**, 723-728 (1978).
30. Elstratidis, A. & Kafatos, F. C. *Cell* **10**, 571-585 (1977).
31. Proudfoot, N. J. *Cell* **10**, 559-570 (1977).
32. Steitz, J. A. & Jakes, K. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4734-4738 (1975).
33. Alberty, H., Raba, M. & Gross, H. J. *Nucleic Acids Res.* **5**, 425-434 (1978).
34. Hagenbuchle, O., Santer, M. & Steitz, J. A. *Cell* **13**, 551-563 (1978).
35. Ziff, E. B. & Evans, R. M. *Cell* **15**, 1463-1475 (1978).

## Design, synthesis and characterisation of a 34-residue polypeptide that interacts with nucleic acids

B. Gutte, M. Däumigen & E. Wittschieber

Max-Planck-Institut für Biophysikalische Chemie, Abteilung Biochemische Kinetik, 34 Göttingen, FRG

*Based on secondary structure prediction rules and model building a neutral artificial 34-residue polypeptide with potential nucleic acid-binding activity was synthesised. This peptide and its covalent dimer showed strong interaction with cytidine phosphates and single-stranded DNA. The dimer had considerable ribonuclease activity with high preference for cleavage at the 3'-end of C.*

BASED on the known X-ray structures of proteins, rules were proposed by various authors (see ref. 1 for review) to describe the relationship between primary and secondary structure of a protein molecule. These rules may be applied to predict the secondary structure of proteins of which only the amino acid sequence is known. An example of this is the suggested succession of the secondary structural elements ( $\alpha$ -helix,  $\beta$ -strand, reverse turn) in *Escherichia coli lac* repressor<sup>2,3</sup>. Proof or disproof of the proposed secondary structure will come from X-ray diffraction studies on crystalline *lac* repressor.