Volume 3 no.12 December 1976 Nucleic Acids Research

Nucleotide sequence of the restriction fragments Hind L and Hind M of SV40 DNA

M.Ysebaert, F. Thys, A.Van de Voorde and W.Fiers

Laboratory of Molecular Biology, University of Ghent, Ledeganckstraat 35, 9000 Gent, Belgium

Received 21 October 1976

ABSTRACT

The nucleotide sequence of the two minor SV40 DNA restriction fragments Hind L and Hind M is here reported. For this purpose we used 5'-^{32}P-labeled DNA fragments either partially digested with snake venom diesterase for analysis by the wandering spot method or partially degraded with the base specific reagents dimethylsulphate or hydrazine for direct sequence analysis on gel. In the former procedure the strands were separated before degradation, while in the latter the strands were separated after modification but before the degradation. Due to the presence of several termination codons, the region Hind L - Hind M cannot be translated in a polypeptide. Also, no initiation codons for protein synthesis are present in this region.

INTRODUCTION

SV40 DNA is cleaved into 13 fragments by the restriction enzyme preparation Hind II + Hind III from Haemophilus influenzae Rd. The 11 major fragments have been ordered by Danna and Nathans (1). The two minor fragments Hind L and Hind M, about 0.50 and 0.40 % of the total genome, were subsequently discovered and localized by Yang et al. (2). Hind L and Hind M are contiguous and map in the late region of the SV40 genome at about 0.75. These fragments could not be cleaved by other available restriction enzymes. We here report their nucleotide sequence.

MATERIALS AND METHODS

Adenosine triphosphate and T_4-polynucleotide kinase were bought from P.L. Biochemicals. γ-^{32}P-ATP was prepared according to Glynn and Chappell (3) using labeled phosphate from NEN.

Bacterial alkaline phosphatase and snake venom phosphodiesterase were obtained from Worthington Biochemical Corp. SV40 DNA was isolated, purified and cleaved into the appropriate restriction fragments as described elsewhere (4,5). A stock-solution of 40 % acrylamide and 2 % bisacrylamide, both from Serva, was purified by passing through a charcoal column, filtered over a 0.22 μ Millipore membrane and stored at 4°C.

The two 5'ends of each restriction fragment were labeled (6) by means of T_4-polynucleotide kinase and γ-^{32}P-ATP, as described by Van de Voorde et al. (7). Instead of the Biogel column purification, the reaction mixture was phenolized in the presence of 10 μg carrier ATP and 10 μg carrier sonicated calf thymus DNA, and precipitated by addition of 2 volumes of ethanol.

The two strands of Hind L and Hind M were separated by a two-dimensional electrophoresis-homochromatography system. After heat denaturation each fragment was applied on a cellulose-acetate strip (60 x 4 cm, from Schleicher & Schüll) and subjected to electrophoresis at pH 3.5 in the presence of 6 M urea. The material was located with a portable Geiger counter and transferred on a DEAE-cellulose thin layer plate (20 x 40 cm, Polygram Cel 300 HR-2/15, from Machery-Nagel & Co). The second dimension was developed at 60°C for 18 hours, using 50 % homomixture α (8) and 50 % homomixture β (9). The separated strands were located by autoradiography, eluted, partially digested with snake venom phosphodiesterase and separated in two dimensions as described previously (7). One sample was also incubated for a very long time until the digestion was complete. It was applied on a polyethyleneimine plate (20 x 20 cm, Polygram Cel 300 PEI, from Machery-Nagel & Co). Formic acid (0.1 % v/v) was allowed to ascend about 2 cm above the origin. The further development was done with 0.2 M LiOH-HCOOH, pH 3.0. In this way the four nucleotides are well separated and the 5'-terminal, ^{32}P-labeled nucleotide of each strand can be identified.

The chemical degradation of DNA with hydrazine (Eastman Kodak) or dimethylsulphate (Aldrich Co) (10) gave no clear-cut results with the single-stranded DNA's derived from homochro-

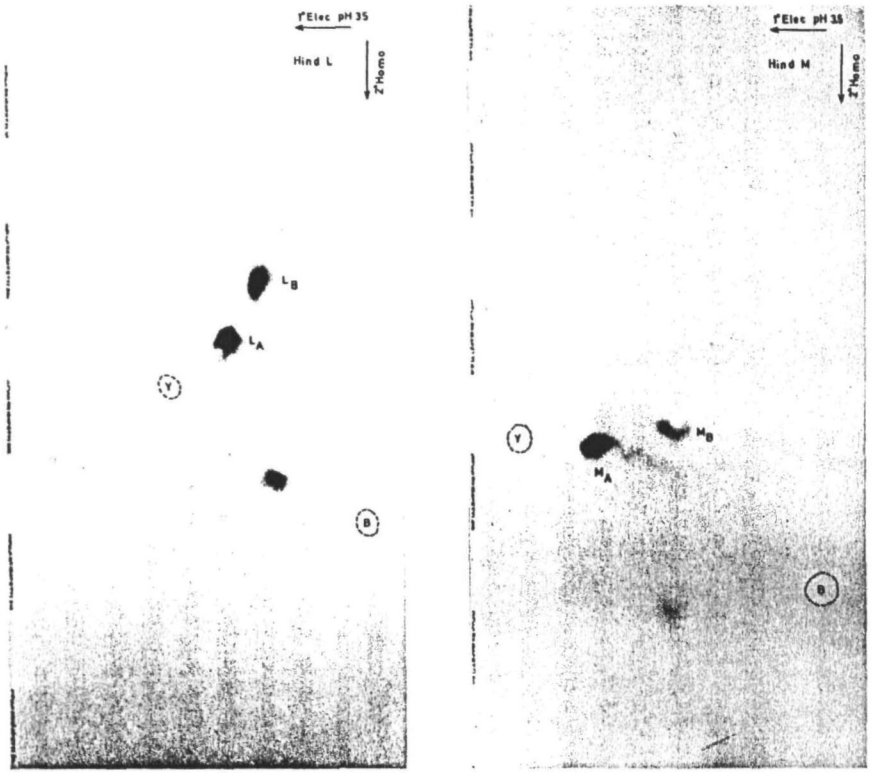


Figure 1 Separation of the two strands of the restriction fragments Hind L and Hind M. The 5'-terminally labeled DNA fragment was heated for two minutes at 100° C in 4 µl 0.02 M Tris-citrate + 6 M urea, pH 8.0. It was quickly chilled and the strands were separated as described under Materials and Methods. B and Y refer to the positions of the xylene cyanol FF (blue) and orange G (yellow) dye markers.

matography. Therefore the chemical modification was carried out on the duplex DNA fragments. The release reactions were preceded by the strand separation (described above) except that the second dimension (homochromatography) was developed at room temperature. Each chemically modified strand was eluted, and further treated for base-specific cleavage as described by Maxam and Gilbert (10). These DNA samples were then analyzed on a 20 % polyacrylamide-urea gel.

RESULTS

The labeling of about 0.1 µg DNA fragment usually resulted in a radioactivity of 5 to 20 x 10^6 disintegrations per minute, depending mostly on the purity of the fragment and on the specific activity of the γ-^{32}P-ATP. Due to the marked difference in base composition between the two strands of Hind L and Hind M, the strand separation presented no difficulties (Fig. 1). Their nucleotide sequence can fairly well be deduced by the wandering spot method and the information, derived from the two strands, is partly complementary (Fig. 2). However, it seemed that a control by another, independent procedure would provide more certainty.

The chemical degradation method, as developed by Maxam and Gilbert (10), gave good results with double strands (5,7) but was not satisfactory with the single strands from these small restriction fragments, even after removal of the RNA by alkaline degradation. The strand separation, however, can be performed after the modification reactions if extreme conditions of pH or temperature are avoided. Some loss of DNA fragment, probably due to premature degradation, and a high RNA concentration after the strand separation are the drawbacks of this modification. By comparing the four nucleotide-specific lanes of each strand on the autoradiograph (Fig. 3 and 4) the nucleotide sequences can again be deduced and is in full agreement with the data from the wandering spot procedure. In the C + T-lane the spots due to degradation of a cytidine residue appear more intense than those corresponding to degradation of a thymidine residue. Since the modification reaction was done on the double strand,

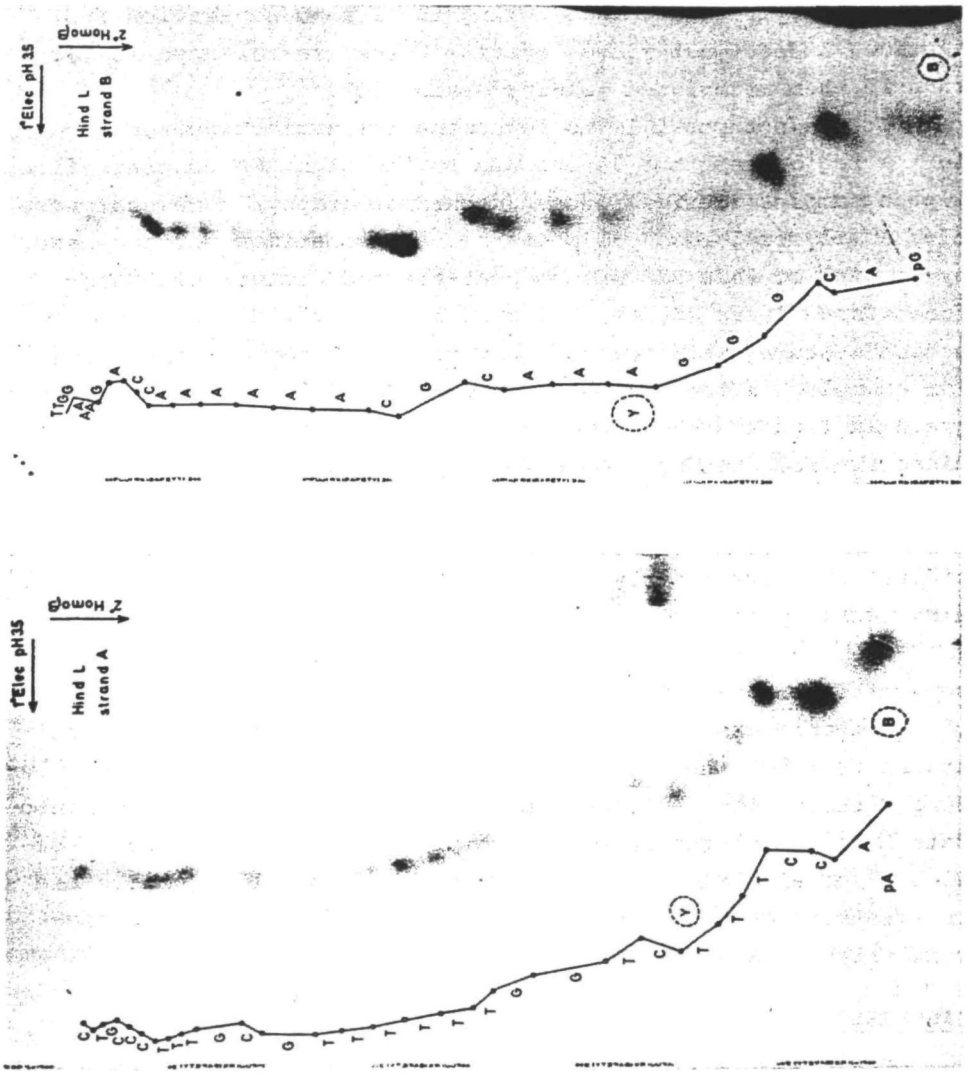
the reaction conditions for the two strands cannot be chosen independently. The 5'-terminal di- and mononucleotides cannot be seen on the autoradiograph. The procedure does not work very well for the desoxyadenylic acid specific breakdown. This might be due to preferential degradation of the strands containing methylated adenines during the strand separation and elution. Indeed, methylated adenine bases are released faster in acid than methylated guanine bases (10).

It was not possible to determine the nucleotide sequence in the region near the 3'-end due to the high RNA concentration in the samples (derived from homochromatography). The sharpness of the polyacrylamide gel pattern could sometimes be increased by removal of this RNA with a heat-treated mixture of DNase-free ribonuclease (5, 11) followed by precipitation of the reaction mixture with ethanol, before the chemical degradation. For example, in the case of strand A from Hind L, the autoradiograph in figure 3 shows that near the 3'-end two or more thymidines are followed by two or more cytidines. In another experiment in which the RNA had been removed as described above, three spots for the thymidine lane and three spots for the cytidine lane were clearly revealed (of course the last ones were also weakly present in the thymidine lane).

The identification of the 5'-terminal nucleotide of the separated strands (Fig. 5) revealed that strand B of restriction fragment Hind L is the only one which starts, with a guanosine residue, the other strand from Hind L and both from Hind M start with a 5'-terminal adenosine residue. The combined data lead unambiguously to the complete nucleotide sequence of Hind L and Hind Hind M (Fig. 6) and the terminal sequences are in agreement with the known recognition sites of the restriction enzymes Hind II (12) and Hpa I (13).

DISCUSSION

The order of the restriction fragments on the genome was found to be Hinc C - Hind L - Hind M - Hind D. The junction between Hinc C and Hind L is the only one which is not cleaved by the Haemophilus parainfluenzae I enzyme (2). Knowing the



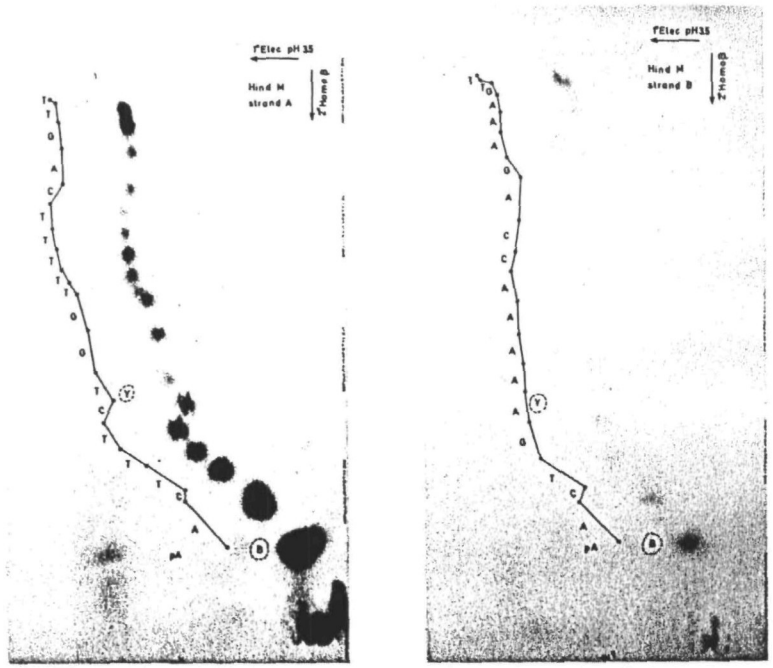


Figure 2 Sequence analysis of restriction fragments Hind L and Hind M by the wandering spot procedure. The single stranded, 5'-terminally labeled DNA fragment was recovered from the homochromatography plate (Fig. 1) and was digested with snake venom phosphodiesterase (7). Samples were removed at different time intervals, pooled and the partial products were separated as described under Materials and Methods. B and Y indicate the positions of the xylene cyanol FF (blue) and orange G (yellow) dye markers. The direction and the extent of the mobility shifts allow a (tentative) identification of the nucleotide removed (note that the mobility shift in the second dimension is larger for a purine than for a pyrimidine residue).

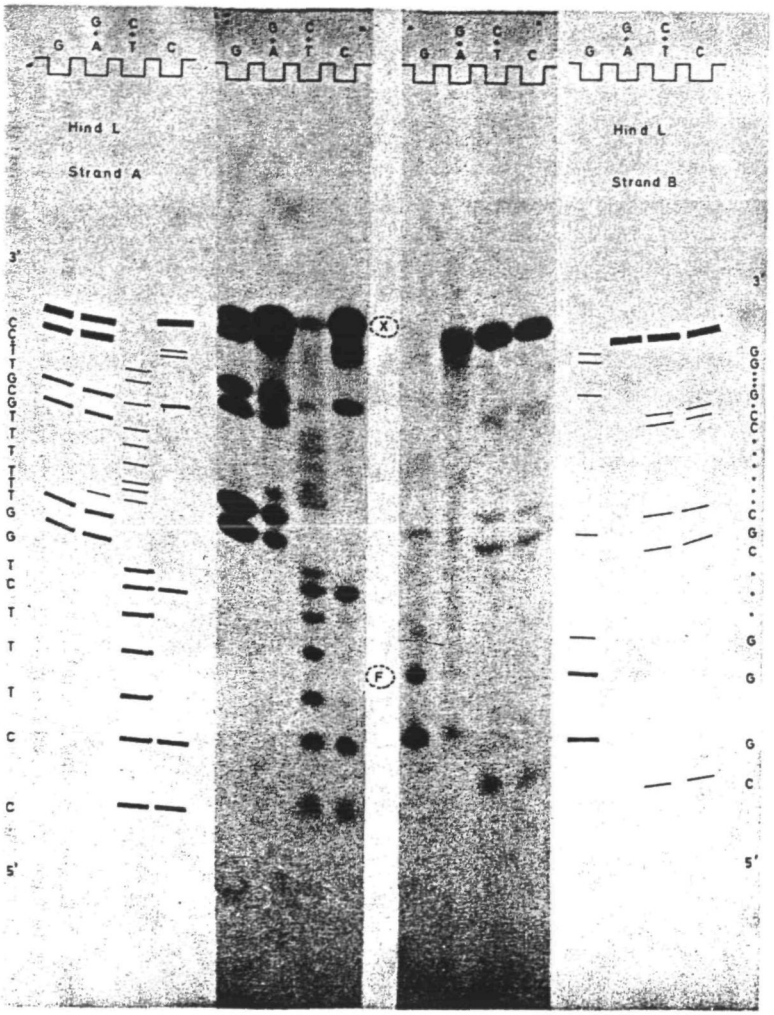


Figure 3 Sequence analysis of restriction fragment Hind L by the chemical degradation procedure. The double stranded, 5'-terminally labeled DNA fragment was mixed with carrier DNA and incubated with dimethylsulphate (G and G + A) or with hydrazine (C and C + T). The reaction, pH 8·0, was stopped by precipitation with ethanol. The DNA was redissolved in 4 µl 0.02M Tris-citrate + 6 M urea, heated for two minutes at 100° C and then quickly chilled. The DNA strands were separated, subjected to base-specific degradation and loaded on a 20% polyacrylamide gel, containing 6 M urea at pH 8·0, as described under Materials and Methods. X and F indicate the positions of the xylene cyanol FF and fluoresceine dye-markers. Modified bases are released and the DNA chain is broken at these positions during an alkaline treatment. This generates 5'-terminally labeled DNA fragments of which the chain length indicates the position of the modified base.

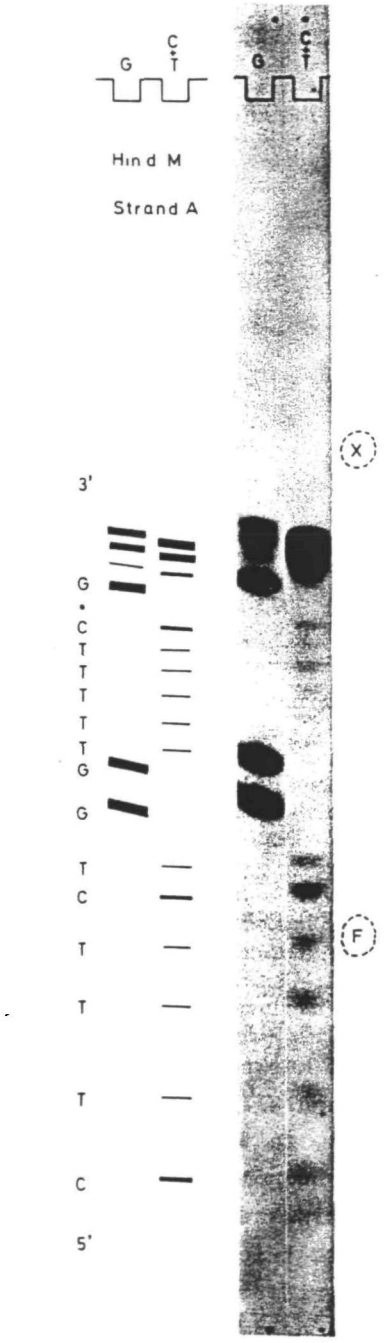


Figure 4 Sequence of the restriction fragment Hind M, strand A, by the chemical degradation procedure. See legend under figure 3.

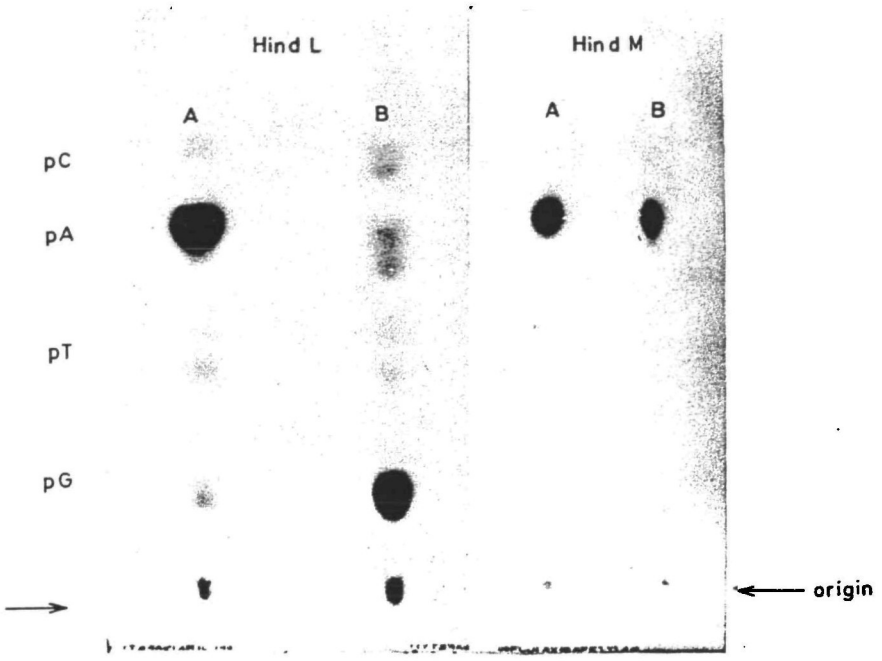


Figure 5 Identification of the 5'-terminal nucleotide of each strand of the restriction fragments Hind L and Hind M. The 5'-terminally labeled DNA strands were totally digested with snake venom phosphodiesterase. The nucleotides were chromatographed on a polyethyleneimine thin layer plate at pH 3.0 as described under Materials and Methods.

base specificities of this enzyme and of restriction enzyme Hind II, this leaves only one possibility for joining Hind C and Hind L. In figure 10 Hind C must be positioned at the left side of Hind L. The lower strand A is transcribed into late messenger RNA. Hind M follows Hind L at the right side but the relative orientation is not known.

In the first orientation all three possible reading frames are blocked by a stop codon. In the second orientation there is one free reading frame although it is also blocked in Hind C (our unpublished nucleotide sequence information). On the other hand there is no ATG nor GTG triplet in these sequences which could be considered as an initiation codon. Therefore, the nucleotide sequence information indicates that the region Hind L - Hind M corresponds to an untranslated part of the SV40 genome. This is supported by the fact that this region is presumably

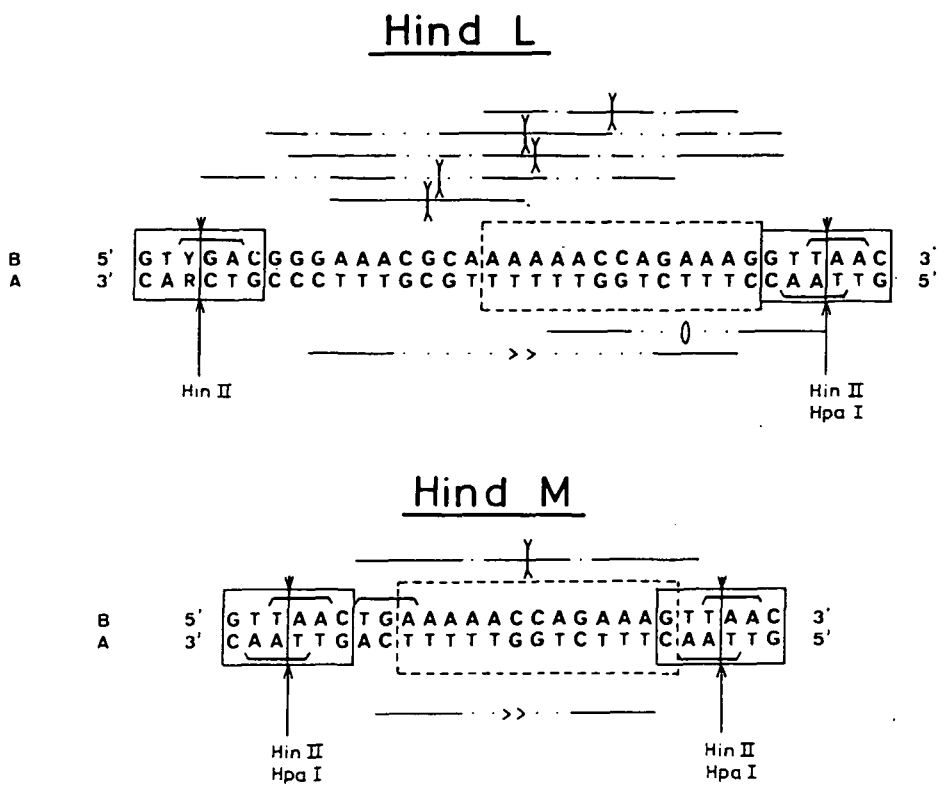


Figure 6 The nucleotide sequence of the restriction fragments Hind L and Hind M. Each DNA fragment is represented with on both sides the recognition sites of the restriction enzymes in full boxes. The dashed boxes indicate a repetition or a palindrome of 13 nucleotides (depending on the relative orientation of Hind M). Symmetrically arranged sequences are represented by full line segments above each fragment. The line segments below the DNA fragments indicate a palindrome (in Hind M) and repetitions (in Hind L and in Hind M). On the left side of Hind L, Y and R stand for pyrimidine and purine respectively as the recognition site of Hind II leaves this question open. The designation B and A refer to the strand separations (Fig. 1).

not essential (14) and by the finding that the three structural viral proteins are coded by a region beyond the Hind M - Hind D junction (15). It is also very striking that in both fragments one strand is particularly A-rich (and the other of course T-rich), making the strand separation very easy. The same orientation of the AT-basepairs also results in the occurrence of many symmetrically arranged sequences (Fig. 6). Even more remarkable is the identical sequence of 13 nucleotides present

in both restriction fragments. Depending on the orientation of Hind M, this might either be a repetition or a palindrome. The probability to find such a symmetry element by chance in a small DNA fragment as presented here is about one in a million. It is quite possible that this region of the genome may be involved in an important biological interaction.

We recently learned that Tu, Roychoudhury and Wu (16) have independently arrived at the same sequence for Hind M.

ACKNOWLEDGEMENT

José Vanderheyden contributed skillful technical assistance. We are grateful to Dr. W. Gilbert and Dr. A. Maxam for valuable informations regarding their sequencing method, and to Dr. K. Kleppe for a generous gift of polynucleotide kinase. This research was supported by a grant from the "Kankerfonds" of the Algemene Spaar- en Lijfrentekas of Belgium. F.T. thanks the I.W.O.N.L. for a fellowship.

REFERENCES

1. Danna, K.J., Sack, G.H. and Nathans, D. (1973) J. Mol. Biol. 78, 363-376.
2. Yang, R., Danna, K.J., Van de Voorde, A. and Fiers, W. (1975) Virology 68, 260-265.
3. Glynn, I.M. and Chappell, J.B.(1964) Biochem. J. 90, 147-149.
4. Yang, R.C.-A., Van de Voorde, A. and Fiers, W. (1976) Eur. J. Biochem. 61, 101-117.
5. Volckaert, G., Contreras, R., Soeda, E., Van de Voorde, A. and Fiers, W., in press.
6. Richardson, C.C. (1965) Proc. Nat. Acad. Sci. U.S.A. 54, 158-165.
7. Van de Voorde, A., Contreras, R., Rogiers, R. and Fiers, W. (1976) Cell, in press.
8. Brownlee, G.G. and Sanger, F. (1969) Eur. J. Biochem. 11, 395-399.
9. Volckaert, G., Min Jou, W. and Fiers, W. (1976) Anal. Biochem. 72, 433-446.
10. Maxam, A. and Gilbert, W. (1975), personal communication.
11. Hiramaru, M., Uchida, T. and Egami, F. (1966) Biochemistry 12, 3055-3063.
12. Kelly, T.J.Jr. and Smith H.O. (1970) J. Mol. Biol. 51, 393-409.
13. Garfin, D.E. and Goodman, H.M. (1974) Biochem. Biophys. Res. Comm. 59, 108-116.
14. Merts, J.E. and Berg, P. (1974) Proc. Nat. Acad. Sci. U.S.A. 71, 4879-4883.

15. Rozenblatt, S., Mulligan, R.C., Gorecki, M., Roberts, B.E. and Rich, A. (1976) Proc. Nat. Acad. Sci. U.S.A., in press.
16. Tu, C.-P.D., Roychoudhury, R. and Wu, R. (1976) Fed. Proc. 35, 1595, abstract nr 1206.