NUCLEOTIDE RECOGNITION SEQUENCE AT THE CLEAVAGE SITE OF *Haemophilus aegyptius* II (*Hae* II) RESTRICTION ENDONUCLEASE

Chen-Pei D. Tu, Ranajit Roychoudhury and Ray Wu

Department of Biochemistry, Molecular and Cell Biology
Cornell University, Ithaca, N.Y. 14853, U.S.A.

Received July 13,1976

SUMMARY: We have determined the nucleotide sequence recognized by the restriction endonuclease *Hae* II from *Haemophilus aegyptius* which cleaves the simian virus 40 (SV40) DNA at a single specific site. By using terminal radioactive labeling of the cleavage site at both the 5' and 3'-ends we have deduced the recognition sequence, 5'...A-G-C-G-C↑pT...3'
3'..Tp↓C-G-C-G--A...5' with elements of a two-fold rotational symmetry. The endonuclease produces staggered ends with protruding 3'-terminated single-strands, four nucleotides in length. In plasmid RSF 2124 DNA, which contains multiple *Hae* II cleavage sites, it was observed that the 5th nucleotide from the 3' terminus is either a pdA or a pdG, indicating alternating recognition sequences.

INTRODUCTION - *Haemophilus aegyptius* contains three restriction endonucleases designated *Hae* I, *Hae* II (1) and *Hae* III (2). Only *Hae* III is well characterized and it has been shown to recognize a tetranucleotide sequence

5'...G-G↓pC-C...3'
3'...C-Cp↑G-G...5' producing an even break at the site of cleavage (2).

The restriction endonuclease *Hae* II cuts the SV40 DNA at a single specific site (1). By using double digestion of the same DNA molecule with two different restriction endonucleases, Roberts *et al* (1) have shown that the *Hae* II recognition sequence is different from that of all other known restriction endonucleases. Only the endonuclease H-1 from *Haemophilus influenzae* strain H-1 produced a gel pattern of DNA bands similar to *Hae* II with λDNA as substrate (1). Takanami and Kojo (3) have found pdC and pdT as the 5' terminal nucleotides at the cleavage site of endonuclease H-1 which indicates that the enzyme probably recognizes two alternate sequences. However, the exact nature of these sequences as well as the recognition sequence of *Hae* II remains unknown(4). In this communication, we report the recognition sequence of this enzyme,

Copyright © 1976 by Academic Press, Inc.
All rights of reproduction in any form reserved.

which was determined after end labeling the single cleavage site in the SV40 DNA, or the multiple cleavage sites in the RSF 2124 DNA, by two independent methods, i.e. labeling 3'-ends with terminal transferase (5) and 5'-ends with polynucleotide kinase (6). The unique labeled ends were isolated by digestion with a second restriction endonuclease and subjected to sequence analysis. By alignment of the resulting nucleotide sequences we have deduced the recognition sequence of *Hae* II endonuclease in these DNA molecules.

MATERIALS AND METHODS - [α-^{32}P]rCTP (specific activities 50-150Ci/mmole) was obtained from New England Nuclear Corporation. [γ-^{32}P]ATP (specific activities 100-200 Ci/mmole) was a product of ICN Pharmaceuticals. Polynucleotide Kinase (specific activity 20,000 units/mg) was obtained from Biogenics Research Corporation. Terminal transferase (specific activity 20,000 units/mg) was isolated as described earlier (7). Restriction endonuclease *Hinc* II (specific activity 9000 units/mg), *Hind* III (specific activity 45,000 units/mg) and *Hae* II (specific activity 20,000 units/mg) were obtained from New England Biolabs. Covalently closed circular SV40 DNA was isolated as described by Wu *et al* (6). The plasmid RSF 2124 DNA (8) was a gift of K. Marians of this laboratory.

Circular Form I SV40 DNA (50 μg) was incubated in 6 mM Tris-HCl, pH 7.8, 6 mM $MgCl_2$, 1 mM dithiothreitol and 10 units of *Hae* II endonuclease in a final volume of 100 μl. After incubation at 37° for 11 hours all of the SV40 DNA molecules were converted to the linear form as revealed by agarose gel electrophoresis (1). The DNA was labeled either at the two 5'-ends using polynucleotide kinase or at the two 3'-ends using terminal transferase as described earlier (6). The labeled DNA was isolated by gel filtration followed by ethanol precipitation. The two labeled ends were separated after a second digestion as follows: The DNA was suspended in 6 mM Tris-HCl, pH 7.8, 6 mM $MgCl_2$, 1 mM dithiothreitol and incubated with 16 units of *Hinc* II and 18 units of *Hind* III at 37° for 7 hours in a final volume of 50 μl. The digested DNA was subjected to polyacrylamide slab gel electrophoresis as described earlier (6). The labeled bands were eluted from the gel and partially digested with pancreatic DNase in order to obtain a homologous series of oligonucleotides from each unique labeled end of DNA (6). In the case of terminal transferase labeled DNA the samples were subjected to alkali and alkaline phosphatase (BAPF grade) treatment (9) before the pancreatic DNase digestion. The digested products were fractionated using two-dimensional electrophoresis-homochromatography (10) and the nucleotide sequence at each labeled end was deduced using calculations for mobility-shift analysis (11).

For determination of *Hae* II recognition sequence in DNA with multiple cleavage site, the plasmid RSF 2124 DNA (50 μg) was digested with 20 units of *Hae* II, 6 mM Tris, pH 7.8, 6 mM $MgCl_2$, 1 mM dithiothreitol at 37° for 22 hours in a final volume of 300 μl. An aliquot (10 μl) of the digested DNA was examined by agarose gel electrophoresis in the presence of ethidium bromide. More than ten fluorescent DNA bands were detectable. The DNA fragments were then labeled with terminal transferase as described earlier (5) and the whole reaction mixture was treated with alkali and alkaline phosphatase (9). After dilution the labeled DNA fragments were absorbed onto a small column of DEAE-cellulose, washed with 0.25M triethylammonium bicarbonate, pH 7.5, and eluted with 2M NaCl, 10 mM Tris, pH 7.8, 100 μM EDTA. After ethanol precipitation

the DNA was partially digested with pancreatic DNase and subjected to two-dimensional electrophoresis-homochromatography (10).

RESULTS - The site of cleavage on SV40 DNA produced by *Hae* II is located between 0.80 - 0.835 unit length (clockwise) from the *Eco* RI site (1). Therefore, the cleavage is made on *Hind* D fragment of SV40 DNA according to the SV40 physical map (12, 13) [see Figure 3]. When the DNA labeled at the *Hae* II cleavage site was digested with *Hind* endonuclease, two labeled fragments designated D_1 and D_2 were obtained (Figure 3). The two labeled duplex DNA fragments, D_1 and D_2 containing only one labeled strand per fragment, were separated by electrophoresis on 5% polyacrylamide slab gel as shown in Figure 3 (on the right-hand side). The other *Hind* fragments were non-radioactive and therefore remain undetected.

When the kinase labeled shorter DNA fragment (D_2) was subjected to two-dimensional mobility-shift analysis, the nucleotide sequence 3'...A-C-T-A-G-T-T-G-T$\overset{*}{p}$ 5' was obtained (Figure 1a). The nucleotide sequence in the comple-

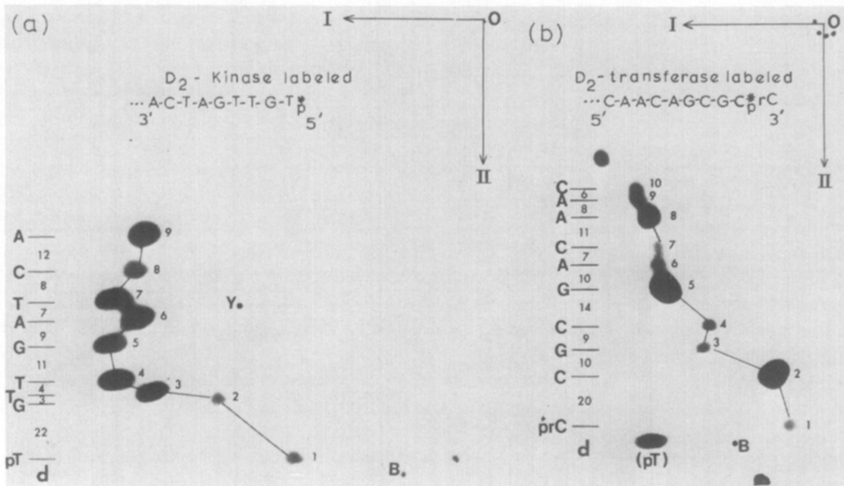


Fig. 1. Nucleotide sequence at the *Hae* II cleavage site in fragment D_2 produced by *Hind* cleavage of SV40 DNA
The Figure shows two-dimensional map of kinase labeled (a) and transferase labeled (b) duplex D_2. The letters on the map indicate the following: O - origin, I - electrophoresis at pH 3.5 on cellulose acetate, II - homochromatography, Y - position of yellow dye (methyl orange), B - position of blue dye (xylene cyanol FF) and the numbers at the left of each map between nucleotides indicate d-value (16) in millimeters.

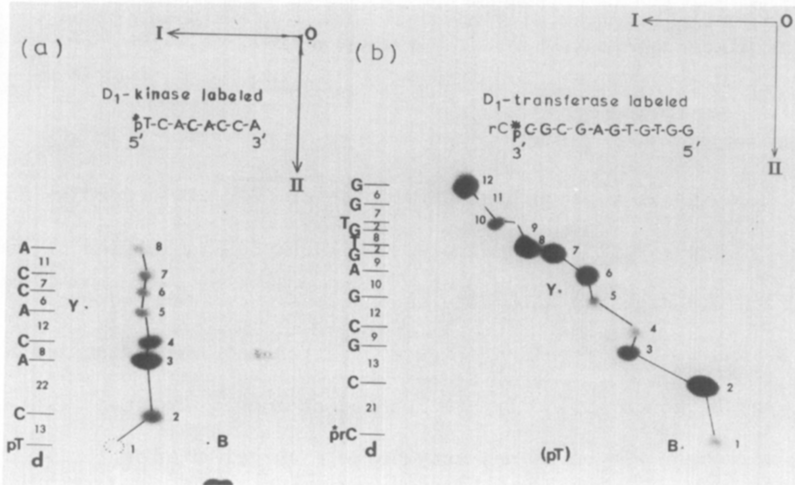


Fig. 2. Nucleotide sequence at the *Hae* II cleavage site in fragment D_1 produced by *Hind* cleavage of SV40 DNA
The figure shows two-dimensional map of kinase labeled (a) and transferase labeled (b) duplex fragment D_1. The other letters on the map are those described in Figure 1.

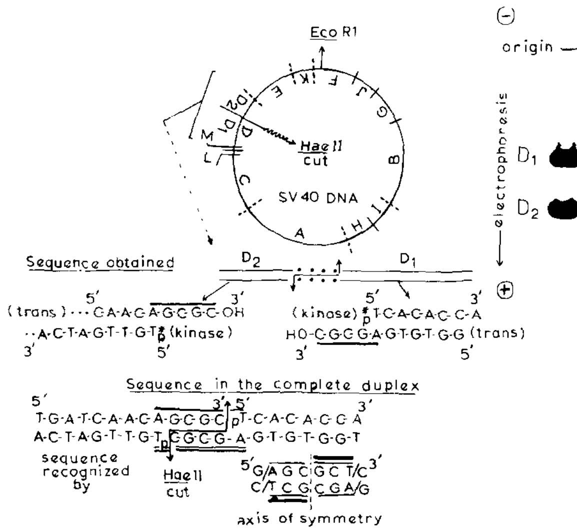
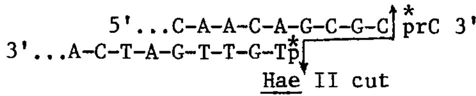
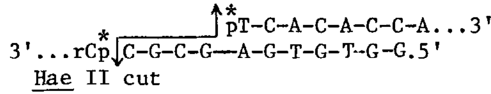


Fig. 3. Schematic presentation of the location, isolation and deduction of a 21 base pair nucleotide sequence at the *Hae* II cleavage site in SV40 DNA

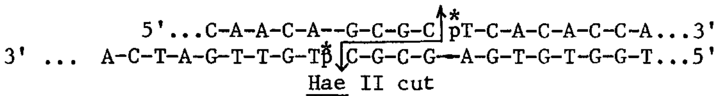
mentary strand of D_2 was obtained as 5'...C-A-A-C-A-G-C-G-CprC 3' (Figure 1b). Therefore, the antiparallel orientation of the two strands yields a duplex:



In the same way, the orientation of the two strands obtained from the longer fragment (D_1) yields a duplex configuration (Figure 2):



The alignment of the two staggered ends of fragments D_2 and D_1 yields a complete duplex nucleotide sequence at the Hae II cleavage site in SV40 DNA as:



To determine whether the above sequence is recognized in every case, or whether there are two sequences recognized by this restriction endonuclease as in endonuclease H-1 (3), the terminal sequences from a mixture of DNA fragments were analyzed. Accordingly, a mixture of about 12 fragments generated by Hae II cleavage of the plasmid RSF 2124 DNA were labeled at the 3'-ends

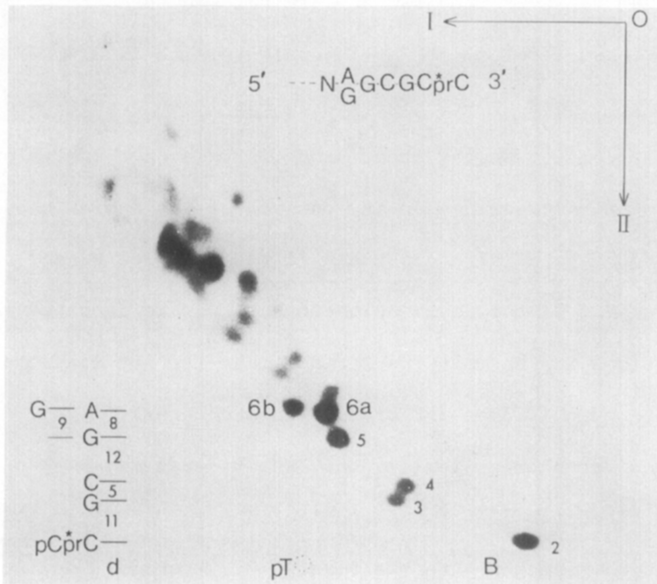
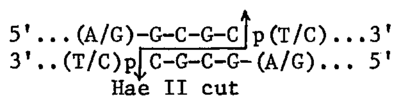


Fig. 4. Two dimensional map of 3' terminal [^{32}P]rCMP labeled mixture of duplex fragments generated by Hae II cleavage of the plasmid RSF 2124 DNA

with [^{32}P]prC by terminal transferase and the sequence analyzed. Figure 4 shows a sequence of 5'..G-C-G-C*pC 3' common to all fragments; moreover, the nucleotide at the 5' side of this common sequence is either a pdA or a pdG. Hence, it appears that Hae II cleaves the DNA to give either 5'..A-G-C-G-C 3' or 5'.. G-G-C-G-C 3', and the general recognition sequence at the cleavage site is:



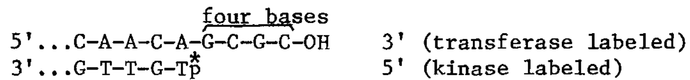
In this way, it became clearly evident that Hae II produces a staggered break with an unusual feature consisting of protruding 3' terminated single-strands, four nucleotides in length.

DISCUSSION - Nucleotide sequence analysis using two independent methods of terminal labeling provides a simple and general means for the determination of the recognition sequence of restriction endonucleases.

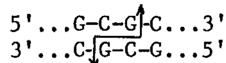
If the cleavage is staggered with protruding 5'-terminal nucleotides, then the 5'-terminal sequences in the separated duplex ends will be complementary to each other. However, the two 5'-terminal sequences from the two 5'-terminal ends (kinase labeled) of Hae II cleavage site in SV40 DNA clearly are not complementary to each other. Therefore, Hae II endonuclease does not produce protruding 5'-ends. On the other hand, if an even break is produced, then the 3' terminal nucleotides in any one end of the duplex should be complementary to the 5' terminal nucleotides in the opposite strand of the same duplex end. When the 3'-terminal nucleotides in D_2 are compared with the 5'-terminal nucleotides in the opposite strand, they are not complementary to each other. Therefore, Hae II does not produce an even break.

When we consider the possibility that the 3'-ends may be protruding, we expect that the first few 3'-terminal nucleotides in one separated end of the duplex should be complementary to the first few 3'-terminal nucleotides in the other end of the duplex. Indeed, we find 5'...A-G-C-G-C-OH 3'
3'..HO-C-G-C-G-A..........5'
the first four 3' terminal nucleotides from the two 3'-ends of the duplex to

be complementary to each other. Furthermore, if the 3'-ends are protruding, then the first few 5' terminal nucleotides in any one end of the duplex must be complementary to the 3' terminal nucleotides in the opposite strand of the same duplex-end, just four nucleotides away from the 3' terminus.



This was found to be the case. Similarly, the D_1-end of the duplex confirms similar observation (see Figure 3). In this way, it is clearly evident that Hae II cut produces four protruding 3' terminal nucleotides. The only other cases of protruding 3' terminal nucleotides are those produced by the endonuclease Hph [with one protruding nucleotide (14)] and Hha 1 [with two protruding nucleotides (15)]. That a hexanucleotide sequence is responsible for Hae II recognition is evident from the following consideration. The endonuclease Hha I recognizes a tetranucleotide sequence,



and cuts at the sites indicated by the arrows producing two 3'-terminal protruding nucleotides (14). This endonuclease cleaves SV40 DNA at two sites. If the Hae-II recognition-site consisted of the same four nucleotides, then it would have cut the SV40 DNA at the two sites also. Since this is not the case and since Hae II makes only one cut on SV40 DNA it must have a longer recognition sequence than that of Hha I. Therefore, we conclude that Hae II recognizes a hexanucleotide sequence, 5'-A-G-C-G-C-T, on SV40 DNA. Similar work with the use of the plasmid RSF 2124 DNA as substrate showed that Hae II recognizes either 5'-A-G-C-G-C-T or 5'-G-G-C-G-C-C.

Barrell and Slocombe (Barrell, B.G. and Slocombe, P.M., personal communication) have independently determined the recognition sequence of Hae II to be 5'-purine-G-C-G-C$^{\downarrow}$-pyrimidine. Furthermore, they found that the purine and the pyrimidine in this hexanucleotide sequence need not be complementary.

Acknowledgement - This work was supported by Research Grant CA-14989 from the National Institutes of Health. This is paper II in a series on "Nucleotide Sequence Analysis of Tumor Virus DNA". Paper I is by Jay, E., Roychoudhury, R. and Wu, R., Biochem. Biophys. Res. Commun. 69, 678-686 (1976).

REFERENCES

1. Roberts, R. J., Breitmeyer, J. B., Tabachnik, N. F. and Myers, P. A. (1975) J. Mol. Biol. 91, 121-123.
2. Murray, K. and Old, R. W. (1974) Progr. Nucl. Acid Res. Mol. Biol. 14, 117-185.
3. Takanami, M. and Kojo, H. (1973) Fed. Eur. Biochem. Soc. Lett. 29, 267-270.
4. Griffin, B. E. and Fried, M. (1976) Methods Cancer Res. 12, 49-86.
5. Roychoudhury, R., Jay, E. and Wu, R. (1976) Nucleic Acid Res. 3, 863-878.
6. Wu, R., Jay, E. and Roychoudhury, R. (1976) Methods Cancer Res. 12, 87-176.
7. Roychoudhury, R. (1972) J. Biol. Chem. 247, 3910-3917.
8. So, M., Gill, R. and Falkow, S. (1976) Mol. and Gen. Genet. 142, 239-249.
9. Kössel, H. and Roychoudhury, R. (1971) Eur. J. Biochem. 22, 271-276.
10. Brownlee, G. G. and Sanger, F. (1969) Eur. J. Biochem. 11, 395-399.
11. Tu, C. D., Jay, E., Bahl, C. P. and Wu, R. (1976) Analyt. Biochem. (in press).
12. Danna, K. J., Sack, G. H. and Nathans, D. (1973) J. Mol. Biol. 78, 363-376.
13. Yang, R., Danna, K., Van de Voorde, A. and Fiers, W. (1975) Virology 68, 260-265.
14. Kleid, D., Zafri, H., Jeffrey, A. and Ptashne, M. (1976) Proc. Nat. Acad. Sci. 73, 293-297.
15. Roberts, R. J., Myers, P. A., Morrison, A. and Murray, K. (1976) J. Mol. Biol. 103, 199-208.
16. Jay, E., Bambara, R., Padmanabhan R. and Wu, R. (1974) Nucl. Acid Res. 1, 331-354.