NUCLEOTIDE SEQUENCE WITH ELEMENTS OF AN UNUSUAL TWO-FOLD ROTATIONAL SYMMETRY IN THE REGION OF ORIGIN OF REPLICATION OF SV40 DNA+

Ernest Jay, Ranajit Roychoudhury and Ray Wu

Section of Biochemistry, Molecular and Cell Biology
Cornell University, Ithaca, New York 14853

Received February 9,1976

ABSTRACT: We have determined the nucleotide sequence at or near the origin of replication of simian virus 40 DNA. This 28-nucleotide-long sequence contains two 8-nucleotide-stretches, complementary to each other, located two nucleotides away from each side of the axis of symmetry. Within each of the 8-nucleotide-stretch is a minor axis of symmetry with 4-nucleotides on each side. This highly symmetrical sequence may be of considerable importance for recognition by a biologically important protein molecule.

In simian virus 40 (SV40), the synthesis of DNA has been shown to be initiated at a specific location in the SV 40 genome (1,2). By comparing the amount of ^{3}H-label in each of the eleven Hind cleaved SV40 DNA fragments obtained from pulse labeled DNA, Danna and Nathans showed that DNA chain growth is bidirectional, originating very close to Hind fragments A/C junction (3). Furthermore, with digestion of partially replicated circular DNA with Eco RI to serve as a reference point, followed by electron microscopic observation, Fareed, Garon and Salzman showed the presence of internal loops composed of two replication forks located at a distance (clockwise) of 0.67 unit length from the Eco RI site (4). This region is very close to the Hind A/C junction which mapped at a distance of 0.655 unit length from the Eco RI site (5).

Recently, a high salt-stable DNA protein complex has been isolated from SV40 infected monkey kidney cell or purified virions (6). High resolution electron microscopy revealed one to three globular protein "knobs" at a specific site in the DNA molecule. Digestion of protein-bound DNA with Eco RI and Hpa II clearly established the location to be in the region of 0.65 - 0.73 unit length of SV40 from the Eco RI site (6). After subjecting the protein-bound DNA

+ This is paper XXVII in a series on "Nucleotide sequence analysis of DNA". Paper XXVI is ref. 8.

to formaldehyde treatment to stablize the complex by a reversible cross-linkage reaction followed by Staphylococcus nuclease digestion to hydrolyze the unprotected region of the DNA, this protein-bound DNA segment has been shown to consist of 25 nucleotide base pairs (7). These interesting observations clearly indicate that a direct DNA sequence analysis of this region may reveal some unusual structural features of the probable replication-site in the SV40 DNA replication. We have, therefore, used a strategy based on the following considerations:

(a) Restriction endonuclease Hind III produces six cleavages on SV40 DNA. One of these sites Hind III B/C junction is identical to the Hind A/C junction.
(b) Hind III can be isolated without traces of contaminating nuclease activities. This is essential for the correct labeling of the DNA terminus.

(i) The enzyme produces cohesive ends with protruding 5'-ends. Thus, the termini of the fragments at Hind A/C junction can be labeled with polynucleotide kinase.

(ii) At the 3'-ends complementary to the protruding 5'-ends, a single [^{32}P] dAMP residue can be added with E. coli DNA polymerase I.

(iii) The 3'-ends in the duplex DNA fragments can also be successfully labeled with ribonucleotides using terminal transferase (8).

Using a combination of all three methods for terminal labeling, we have analyzed the nucleotide sequence at the Hind fragments A/C junction of SV 40 DNA. In this paper we report the sequence of a total of 28 nucleotide base pairs possessing an unusual structural feature which might be located at or in the vicinity of the origin of replication of SV40 DNA.

MATERIALS AND METHODS

[α-^{32}P]rNTPs and [α-^{32}P]dNTPs (specific activities 50-150 Ci/mmole) were obtained from New England Nuclear Corporation. [γ-^{32}P]rATP was a product of ICN Pharmaceuticals. Terminal transferase (specific activity 20,000 units/mg) was isolated as described earlier (9). Polynucleotide kinase (specific activity 20,000 units/mg) was obtained from Biogenics Research Corporation. E. coli DNA polymerase I (specific activity 5000 units/mg with calf thymus DNA as template-primer) was isolated according to Jovin et al (10). Restriction endonucloease Hind(a mixture of Hind II and Hind III) was isolated according to Smith and Wilcox (11) and further purified to obtain Hind III completely free from Hind II and other contaminating nucleases. The endonuclease Hae III was purified according to Roberts et al (12). Covalently closed circular SV40 DNA was isolated as described by Wu et al (13). Hind III fragments of SV40 DNA

were labeled at the 3'-ends with [α-32P]rCTP and the terminal transferase[a] as described earlier (8). The 3'-ends of Hind III fragments were also labeled with a single dAMP residue using DNA polymerase I (13). The 5'-ends were labeled using polynucleotide kinase (13). The labeled duplex DNA fragments isolated by gel filtration were digested with Hae III and the unique labeled ends were separated by polyacrylamide slab gel electrophoresis (8). The labeled bands were eluted from the gel and partially digested with pancreatic DNAse in order to obtain a homologous series of oligonucleotides (13). These were subjected to two-dimensional electrophoresis-homochromatography (14) followed by the improved method for sequence analysis by mobility shifts (15).

RESULTS

When terminally labeled Hind III fragments (6 fragments: A,B,C,D,E,F) are redigested with Hae III, the fragment D remains undigested due to the absence of a Hae III site. Thus, eleven labeled bands are produced after redigestion (8). Each of these bands (except Hind III-D) now contains only one unique labeled end. These labeled bands were separated by polyacrylamide gel electrophoresis. The band number 4 from Hind III-B and the band number 10 from Hind III-C (numbered according to their order of mobility) originate from Hind III fragments B/C junction and correspond to Hind fragments A/C junction (5). [See SV40 physical map in Figure 3, details will be published elsewhere].

When the band number 10 labeled at the 5'-end was partially digested by pancreatic DNase and subjected to two-dimensional fractional and mobility shift analysis, the nucleotide sequence[b] $\overset{*}{p}$-A-G-C-T-T-T-T-T-G-C was obtained (Fig. 1-I). Since the nucleotide sequence at the Hind III cleavage site is known (16) to consist of a staggered duplex with protruding 5'-end, $\begin{matrix} 3'...T\text{-}T\text{-}C\text{-}G\text{-}Ap\ 5' \\ \quad ...A\text{-}OH \end{matrix}$, only one [^{32}P]dAMP residue was added during repair-synthesis (13) in the presence of DNA polymerase I and [α-^{32}P]dATP alone (fig.1-II). Partial digestion of this labeled end followed by mobility shift analysis gave the sequence 5' G-G-C-T-T-T-T-G-C-A-A-A-A-$\overset{*}{p}$A (Fig. 1-II). When the 3' terminus was labeled with [α-^{32}P]rCTP and terminal transferase, the sequence 5' G-C-T-T-T-T-G-C-A-A-A-A-rC was obtained (Fig. I-III). In this way, the sequence at the 3' terminus of fragment number

(a) Terminal transferase labeled ends were treated with alkali and phosphatase before the pancreatic DNase digestion.

(b) $\overset{*}{p}$ denotes a [^{32}P]phosphate. All the sequences presented in this paper are those of deoxynucleotides, the prefix d was omitted for simplicity.

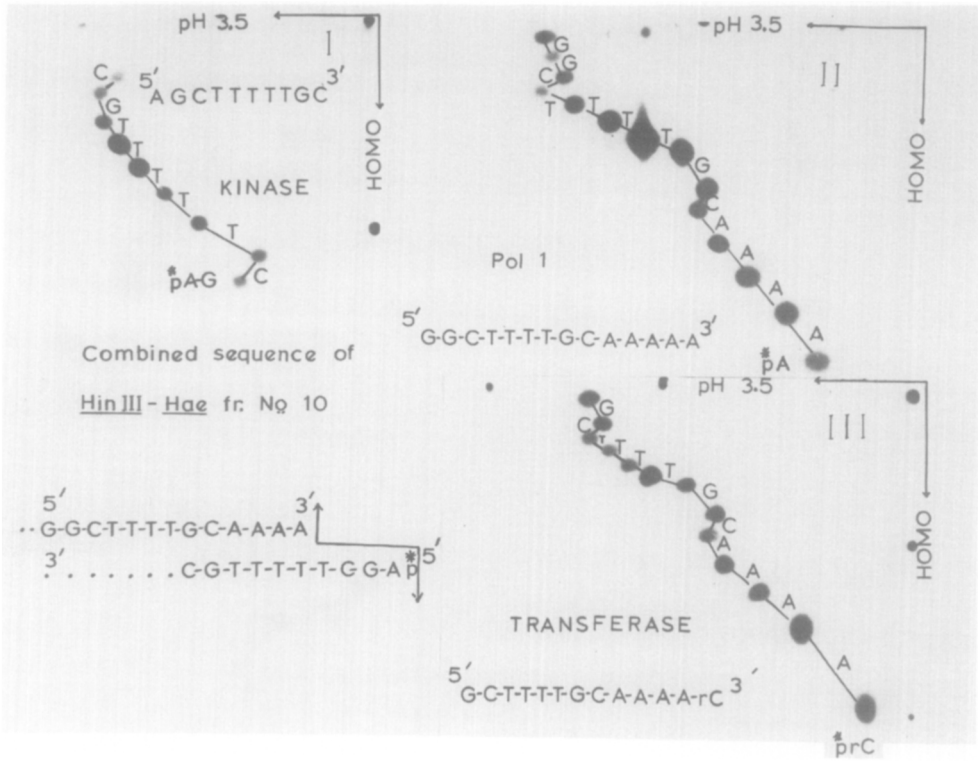


Figure 1. Nucleotide sequences at the duplex right-end of Hind III-C fragment.

Hind III cleaved duplex DNA fragments labeled at both the 5'-ends (kinase labeled) or both the 3'-ends (Pol 1 or transferase labeled) were digested with Hae III and the unique labeled ends were separated by polyacrylamide gel electrophoresis. The band number 10 originating from the right-end of Hind III-C fragments (see Fig. 3) was eluted and partially digested with pancreatic DNase in order to obtain one homologous series of oligonucleotides. These were subjected to two dimensional separation by electrophoresis on cellulose acetate at pH 3.5 in the first dimension and homochromatography (14) in the second dimension. The figure shows two-dimensional map of kinase labeled (I), DNA polymerase labeled (II) and terminal transferase labeled (III) band 10.

10 is fully confirmed. Because of the antiparallel orientation of the DNA double helix we may deduce the sequence at the Hind A/C junction for the Hind-C or Hind III-C fragment as:

5' G-G-C-T-T-T-T-G-C-A-A-A-A- 3' (Pol 1 or transferase labeled)
3' C-G-T-T-T-T-T-C-G-A*p 5' (kinase labeled) .....(a)

In the same way, the band number 4 labeled with polynucleotide kinase, DNA

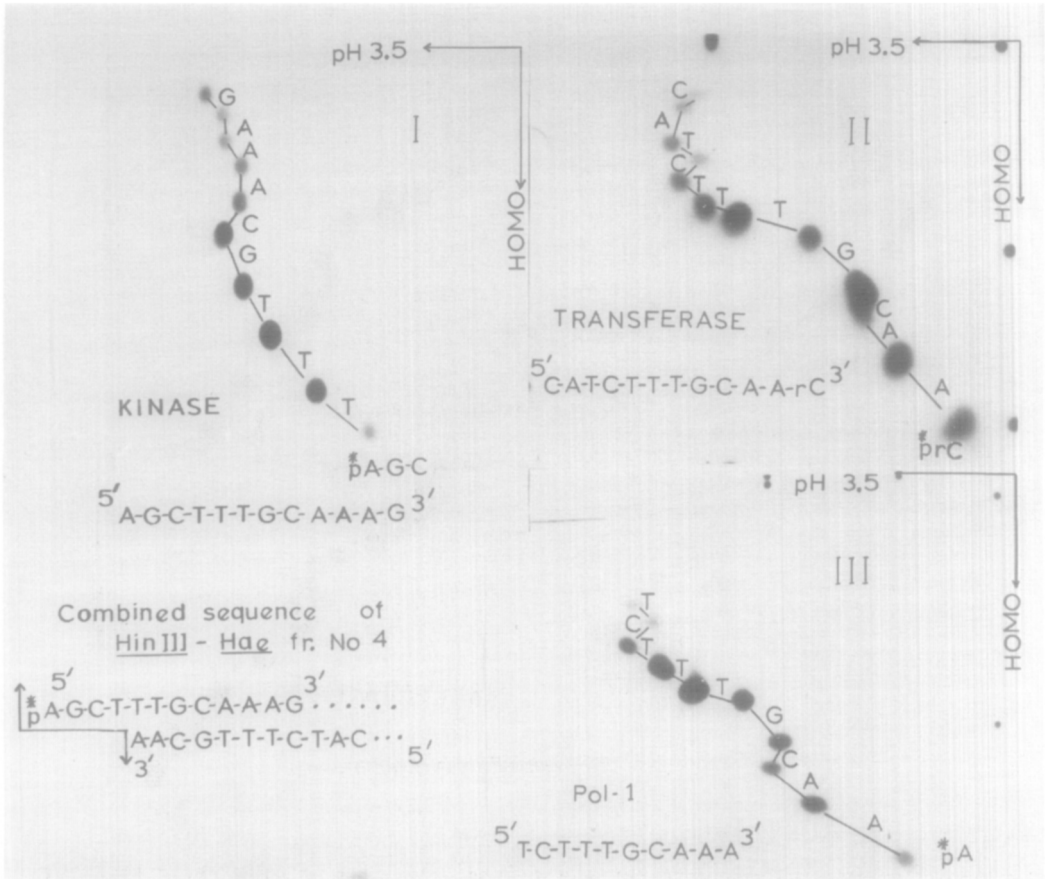


Figure 2. Nucleotide sequences at the duplex left-end of Hind III-B fragment.

The labeled fragment from the left-end of Hind III-B fragment (band number 4, see Fig. 3) was isolated and processed as described in Fig. 1. The figure shows two-dimensional map of kinase labeled (I), terminal transferase labeled (II) and DNA polymerase labeled (III) band 4.

polymerase and terminal transferase gave the sequences 5' *pA-G-C-T-T-T-G-C-A-A-A-G, 5' T-C-T-T-T-G-C-A-A-A and 5' C-A-T-C-T-T-T-G-C-A-A-rC respectively (Fig. 2). All the sequences are deduced with the help of mobility calculations (15) and not simply by visual inspection. From these, the sequence at the Hind A/C junction for the Hind-A or Hind III-B fragment can be deduced as:

5' *pA-G-C-T-T-T-G-C-A-A-A-G 3' (kinase labeled)

3' A-A-C-G-T-T-T-C-T-A-C 5' (transferase labeled).....(b)

Therefore, the alignment of the two Hind III staggered ends [(a) and (b)] would yield a sequence:

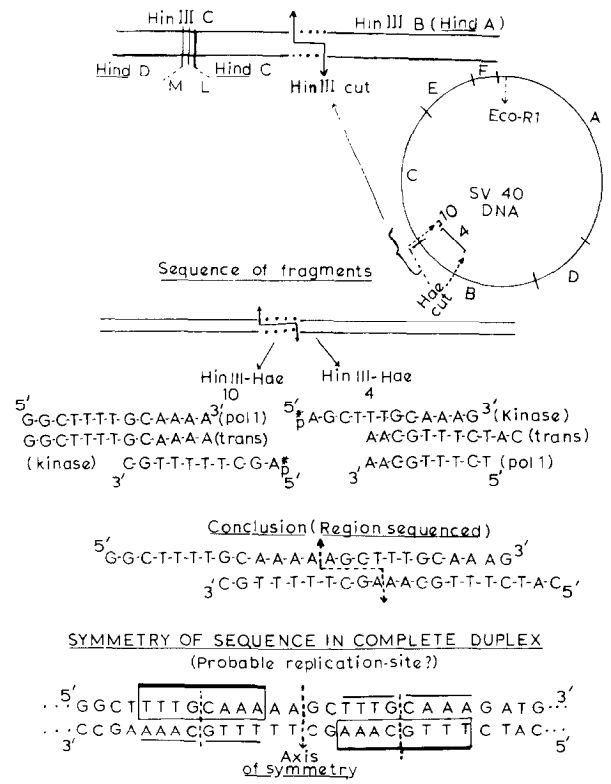
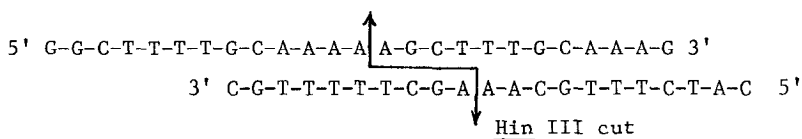


Figure 3. Schematic presentation of the location, isolation and deduction of a 28 base-pair nucleotide sequence with elements of two-fold rotational symmetry.



In this way we have deduced a total of 28 nucleotide base pairs (in the completed duplex form) at or near the origin of replication of the SV40 chromosome (see Fig. 3).

DISCUSSION

In recent years, protein-DNA interactions are increasingly recognized to play a vital role in the mechanisms for the biological control of gene expression. The interaction of spleen DNase and restriction endonucleases reveal nucleotide sequences at the site of cleavage to consist of an axis of two-fold

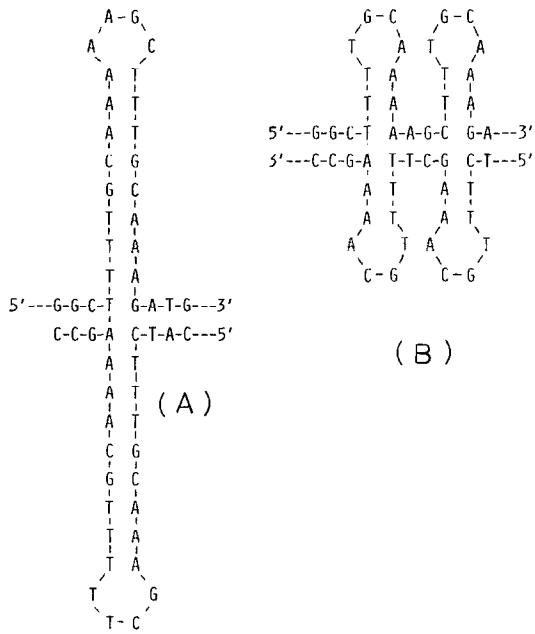


Figure 4. Probable transient folded configurations of the symmetrical sequence for recognition by a biologically active protein.

(A) and (B) are probable structures which may participate in protein-DNA interaction.

rotational symmetry. Such symmetrical sequences (complementary palindromes) have been found at the interaction sites for several restriction endonucleases, comprising tetra, penta or hexanucleotides. For more important biological control functions larger nucleotide sequences are expected to be required for unique recognition; such as the ter function at the cohesive ends of λDNA (17,18), lac operator (19-21), λ operator (22), fd DNA promoter (23), tyrosine tRNA gene promoter (24), lac promoter (20), E. coli RNA polymerase binding site in SV40 DNA (25,26) and T7 promoter (27).

The most remarkable feature of the sequence of 28 nucleotide base pairs shown in Fig. 3 concerns the presence of not only a highly ordered two-fold rotational symmetry around the axis of symmetry, but also the presence of two identical stretches of 8 continuous nucleotides, just two nucleotides away

from the major axis of symmetry. Besides the major axis of symmetry of the palindrome a minor axis with two-fold rotational symmetry exists within each of the two 8-nucleotide-long stretches. The two 5' T-T-T-G as well as the two 5' C-A-A-A sequences occur diagonally around each of the two minor axes. Similarly, the sequence 5' T-T-T-G-C-A-A-A which occurs twice on the same strand also occurs in the complementary strand. This unusually high degree of symmetry may allow the molecule to assume (disregarding stability) two different folded structures (Fig. 4) which may be of some biological importance in relation to protein-DNA interaction. However, specific interaction between protein and DNA molecules possessing two-fold rotational symmetries may occur without the need of any looped-out structure.

If N is the total number of nucleotides examined on one side of the axis of symmetry, B is the number of these nucleotides which are symmetrical, then the probability of any symmetrical sequence (B/N = 8/10) occurring by chance (28) in any 20-nucleotide long segment is very low (probability = 0.0416%). The fact that there are additional symmetries within each of the two 8-nucleotide-long stretches makes this sequence rather unique in the entire 5000-nucleotide-long SV40 DNA.

In the absence of direct biochemical evidence it is difficult to assign a definite function for this interesting region of DNA. However, the highly ordered symmetry in this stretch of 28 nucleotide base pairs, with its possibility of interacting with 2 or 4 subunits of protein and its location in the SV40 genome considered in the context of a protein binding site in this region (6), strongly suggests that this region of ordered symmetry is probably related to the bi-directional replication of SV40 chromosome.

Acknowledgement - This work was supported by Research Grants CA-14989 from the National Institutes of Health, and BMS 73-01859 A02 from the National Science Foundation.

REFERENCES

1. Nathans, D. and Danna, K.J. (1972) Nature, New Biol. 236, 200-202.
2. Thoren, M., Sebring, E.D., and Salzman, N.P. (1972) J. Virol. 10, 462-468.
3. Danna, K.J. and Nathans, D. (1972) Proc. Nat. Acad. Sci. U.S. 69, 3097-3100.
4. Fareed, G.C., Garon, C.F. and Salzman, N.P. (1972) J. Virol. 10, 484-491.

5. Danna, K.J., Sack, G.H. and Nathans, D. (1973) J. Mol. Biol. 78, 363-376.
6. Griffith, J., Dieckmann, M. and Berg, P. (1975) J. Virol. 15, 167-172.
7. McCann, P.P. and Martin, R.G. (1975) Biochim. Biophys. Acta 395, 280-283.
8. Roychoudhury, R., Jay, E. and Wu, R. (1976) Nucleic Acid. Res. 3, 101-116.
9. Roychoudhury, R. (1972) J. Biol. Chem. 247, 3910-3917.
10. Jovin, T.M., Englund, P.T. and Bertsch, L.L. (1969) J. Biol. Chem. 244, 2996-3008.
11. Smith, H.O. and Wilcox, K.W. (1970) J. Mol. Biol. 51, 379-391.
12. Roberts, R.J., Breitmeyer, J.B., Tabachnik, N.F. and Myers, P.A. (1975) J. Mol. Biol. 91, 121-123.
13. Wu, R., Jay, E. and Roychoudhyry, R. (1976) Methods Cancer Res. 12 (in press).
14. Brownlee, G.G., Sanger, F. and Barrell, B.G. (1968) J. Mol. Biol. 34, 379-412.
15. Tu, C.D., Jay, E., Bahl, C.P. and Wu, R. (1976) Analyt. Biochem. (in press).
16. Old, R., Murray, K. and Roizes, G. (1975) J. Mol. Biol. 92, 331-339.
17. Weigel, R.H., Englund, P.T., Murray, K. and Old, R.W. (1973) Proc. Nat. Acad. Sci. U.S. 70, 1151-1155.
18. Ghangas, G.S., Jay, E., Bambara, R. and Wu, R. (1973) Biochem. Biophys. Res. Comm. 54, 998-1007.
19. Gilbert, W. and Maxam, A. (1973) Proc. Nat. Acad. Sci. U.S. 70, 3581-3584.
20. Dickson, R.C., Abelson, J., Barnes, W. and Reznikoff, W. S. (1975) Science 187, 27-35.
21. Bahl, C.P., Wu, R., Itakura, K., Katagiri, N. and Narang, S.A. (1976) Proc. Nat. Acad. Sci. U.S. 73, 91-94.
22. Maniatis, T., Ptashne, M., Barrell, B.G. and Donelson, J. (1974) Nature 250, 394-397.
23. Schaller, H., Gray, C. and Herrmann, K. (1975) Proc. Nat. Acad. Sci. U.S. 72, 737-741.
24. Sekiya, T., Van Ormondt, H. and Khorana, H.G. (1975) J. Biol. Chem. 250, 1087-1098.
25. Zain, B.S., Weissman, S.M., Dhar, R. and Pan, J. (1974) Nucl. Acid Res. 1, 577-594.
26. Dhar, R., Weissman, S.M., Zain, B.S., Pan, J. and Lewis, A.M., Jr. (1974) Nucl. Acid. Res. 1, 595-614.
27. Pribnow, D. (1975) Proc. Nat. Acad. Sci. U.S. 72, 784-788.
28. Dykes, G., Bambara, R., Marians, K. and Wu, R. (1975) Nucl. Acid Res. 2, 327-346.