Volume 1 number 4 April 1974 Nucleic Acids Research

The nucleotide sequence preceding an RNA polymerase initiation site on SV40 DNA. Part 2. The sequence of the early strand transcript

R.Dhar, S.M.Weissman, B.S.Zain, J.Pan, and A.M.Lewis, jun.*

Departments of Human Genetics, Medicine, and Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, Connecticut, USA

Received 18 February 1974

ABSTRACT

The nucleotide sequence of the RNA transcript from tne "early" (E) strand of SV40 DNA immediately preceding the preferred E. coli RNA polymerase start site is G-(A-A-A-C, -A-U-)-A-A-A-A-U-G-A-A-U-G-C-A-A-U-U-G-U-U-G-U-U-G-U-U-A-A-C-U-U-G-U-U-U-A-U-U-G-C-A-G-C-U-U-A-U-A-A-U-G-G-U-U-A-C-Ap. The last nucleotide of the sequence is the first nucleotide transcribed by E. coli RNA polymerase from the "E" strand. The DNA template contains a palindrome of 17 residues that includes the Hemophilus influenza restriction endonuclease cleavage site G-T-T-A-A-Cp. The DNA which gives this transcript lies very close to one end of SV40 DNA segment in the Adeno-SV40 hybrid virus Ad2+ND3 and appears to contain sufficient untranscribed information to specify the E. coli RNA polymerase start.

INTRODUCTION

SV40 DNA is a superhelical closed circular DNA, with molecular weight of 3.5×10^6 daltons. Westphal(1), showed that transcription of Form I SV40 DNA by E. coli RNA nucleotidyl transferase (RNA polymerase E.C. 2. 7. 7. 6) is assymmetric, and transcription is from the "early" strand (E-strand). He also found that some of the transcripts obtained were larger than the unit length of the SV40 genome. We have reported (2) that there is a preferred site of initiation by E. coli RNA polymerase holoenzyme located in the SV40 DNA segment con-

Abbreviations used: SV40 - Simian Virus 40. Hin - Hemophilus influenza. "E" strand - That strand of SV40 DNA whose transcript accumulates in the cytoplasm of infected cells prior to replication of viral DNA. "L" strand - The strand of SV40 DNA complementary to the "E" strand.

tained in the Adeno-SV40 hybrid virus, Ad2+ND3 (3). Because E. coli RNA polymerase can traverse the circle and make transcripts larger than the unit length of SV40 DNA, we were able to obtain nucleotide sequences preceding and overlapping the preferred site of initiation by hybridizing SV40 complementary RNA to Ad2+ND3 DNA. Simple and reproducible sets of oligonucleotides were obtained by digestion of RNA which had been transcribed from SV40 DNA and annealed to Ad2+ ND3 DNA, or transcribed from Ad2+ND3 DNA and annealed to SV40 DNA. Since the initiation of transcription lies over 60 nucleotides from the end of the SV40 DNA segment present within Ad2+ND3, sequence analysis of the annealed RNA has given us the structure of a substantial region of DNA immediately preceding the start site.

MATERIALS AND METHODS

Most of these were described in the preceding report(4). For DNA isolation, Ad2+ND3 virus was propagated in suspension cultures of the KB line of human epidermal cells and purified as described previously (2). The infected cells were frozen and thawed three times and treated with deoxycholate (1%), trypsin (0.01%), pancreatic RNase (10μ/ml) at 37°C for 30 minutes. Virus was pelleted by high speed centrifugation into a cushion of CsCl solution (density 1.43) and purified by equilibrium density centrifugation in CsCl solution (density 1.34). The salt was removed by dialysis and DNA extracted by disruption of the virus with pronase and .5% sodium dodecyl sulphate, followed by phenol extraction. After the treatment two volumes of alcohol were added to the aqueous layer and the DNA recovered by spooling and dissolved in .015 M sodium chloride, .0015 M sodium citrate (pH 7).

RESULTS

Labelled SV40 complementary (cSV40) RNA was prepared by transcription of Form I SV40 DNA with E. coli RNA polymerase holoenzyme. The transcript was hybridized to Ad2+ND3 DNA immobilized on the filters and eluted at 100°C in the presence of 100 μg of carrier t-RNA. Oligonucleotide maps of T_1 RNase and pancreatic RNase digests were prepared

from RNA, labelled separately with each of the four α-^{32}P labelled ribonucleoside triphosphates. These maps showed discrete and reproducible patterns of oligonucleotides (Fig. 1 and Fig. 2).

The nucleotide sequence previously published (5) contained all the T_1 RNase products of the SV40 RNA complementary to Ad2+ND3 DNA except those listed in Table 1. Analysis by further digestion with pancreatic RNase was sufficient to determine the sequence of many of the oligonucleotides. To deduce the sequence of T_1GE 114, 553 and 8, we used U_2 RNase (6) to cleave adjacent to internal adenylic acid residue (Table 2).

Table 3 presents the analysis of pancreatic RNase digestion products of that segment of SV40 DNA which preceded the start of E. coli RNA polymerase and was contained within Ad2+ND3. The sequence of these products was determined by digestion with T_1 RNase.

To deduce the order of T_1 RNase digestion products limited digestion of the RNA was done with "C" RNase (7, 8), carboxymethylated pancreatic RNase (9, 10), and spleen acid RNase (11).

Table 4 shows the analysis of those products of "C" RNase digestion of cSV40 RNA hybridized to Ad2+ND3 DNA which came from the sequence preceding the E. coli RNA polymerase start. Further information about the sequence was derived from the partial digestion products obtained with carboxymethylated pancreatic RNase (Table 5) and spleen acid RNase (Table 6). Here also analysis are shown only for those sequences which preceded the E. coli RNA polymerase start and were contained within Ad 2+ ND3.

The information obtained from these partial digestions was enough to order completely these T_1 RNase digestion products (Fig. 3). The initial sequence of the E. coli RNA polymerase transcript is pppA-A-A-U-A-A-A-G-Cp(5). This base sequence is found in two T_1RNase products (oligonucleotides T_1GE 8 and 2) of SV40 cRNA which had been annealed to Ad2+ND3 DNA. However oligonucleotide 2 was transcribed in its

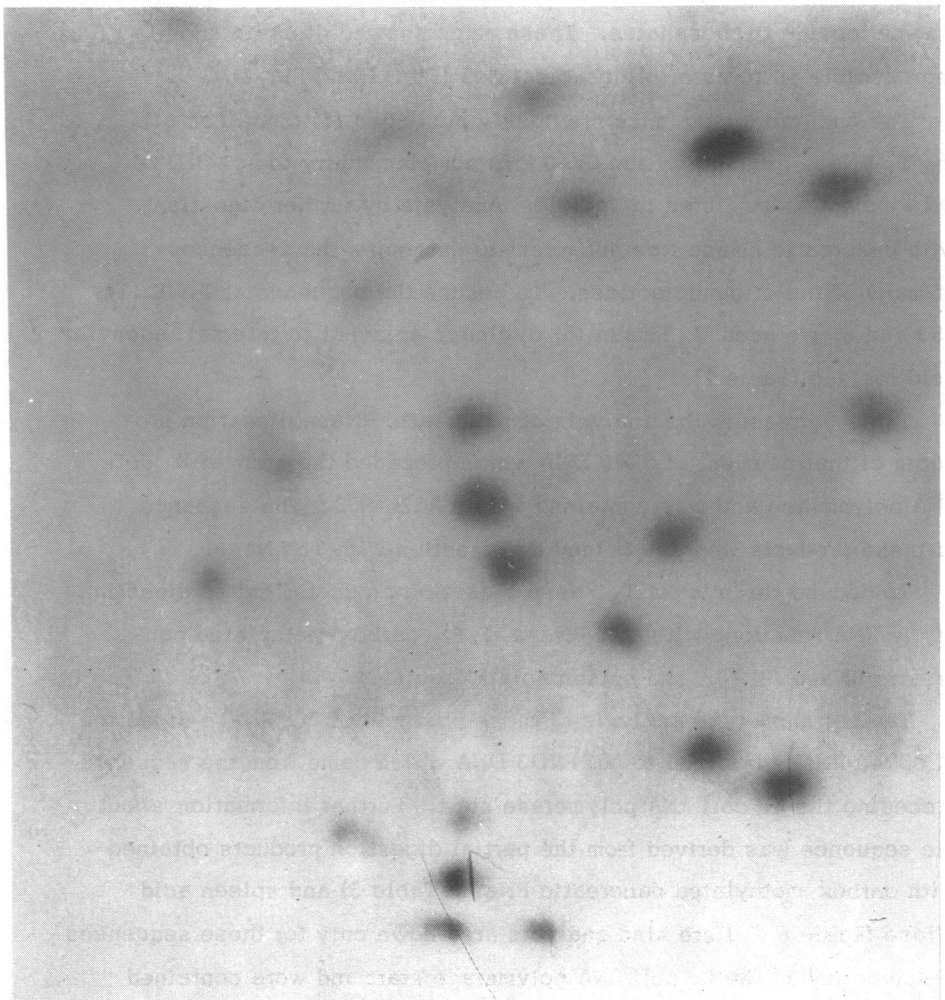
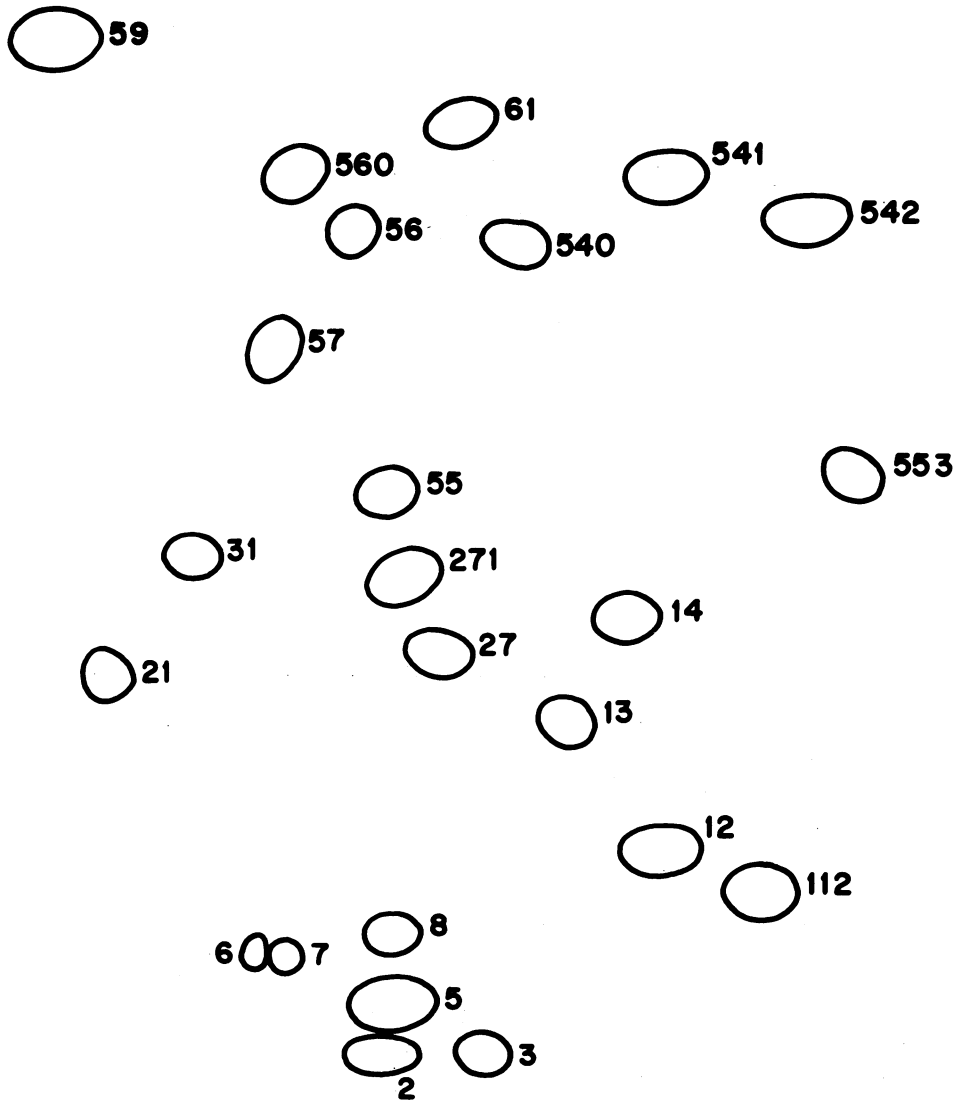


Figure 1

A. Autoradiograph of a two dimensional electrophoretic and chromatographic separation of oligonucleotides produced by T_1 RNase digestion of SV40 complementary RNA annealed to Ad2+ND3 DNA. Electrophoresis was from left to right on Cellogel in 7M urea at pH3.5 and chromatography from bottom upwards on a DEAE cellulose thin layer chromatography plate with "home B" solution (17). The RNA was synthesized on an SV40 DNA template in the presence of α-^{32}P UTP(36.6 Ci/m mol) for 30 minutes at 37°C, and annealed to Ad2+ND3 DNA, eluted and digested with T_1 RNase as described elsewhere(2).



B. Schematic sketch of autoradiograph of 1(A).

TABLE 1

PRODUCTS OF PANCREATIC RNASE TREATMENT OF OLIGONUCLEOTIDES FROM SV40 cRNA WHICH HAD BEEN HYBRIDIZED TO Ad2+ND3 DNA ELUTED, AND DIGESTED WITH T_1 RNASE

Radioactive Precursor Oligonucleotide	Pancreatic RNase digestion products α-^{32}P ATP	α-^{32}P CTP	α-^{32}P GTP	α-^{32}P UTP	Sequence deduced
T_1GE 7	AAAC, AU AAAAU, G	AAAC	AAAAU	AAAAU, AU	(AAAC, AU)AAAAUG(A)
T_1GE 57	AAU	G	AAU	AAU	AAUG(C)
T_1GE 55	C, AAU		U	AAU, G	CAAUUG(U)
T_1GE 541			U	U, G	UUG(U)
T_1GE 114	AAC, U	AAC	U	AAC, U, G	U(AAC, U, U)UG(U)
T_1GE 553	U	G	U	AU, U	U(AU, U, U)UG(C)
T_1GE 47	C	AG	AG		CAG(C)
T_1GE 13	AAU, AU, U		AAU, G	AAU, AU, C, U	CUUAUAAUG
T_1GE 59				G	G(U)
T_1GE 8	AC, AAAG, AAAU, U	AAAG, AC	AAAU, U	AAAU, U	(UU, AC, AAAU)AAAG(C)
T_1GE 31	AAU, C	AG	AG	AAU	CAAUAG(C)

Table 1

The RNA sample was digested with T_1RNase. The resulting oligonucleotides were separated by electrophoresis on Cellogel strips at pH 3.5 and homochromatography (17) on thin layer plates. Individual products were eluted, digested with pancreatic RNase. Then digestion products were analyzed on DEAE cellulose paper by electrophoresis at pH 3.5, and the resulting products identified by their mobility. Brackets indicate the nearest neighbor base. In this and subsequent tables, symbols for phosphate residues are omitted and each letter G C U or A stands for the 3' phosphoryl nucleoside.

entirety in pulse synthesis and lies beyond the start site of RNA polymerase while T_1GE 8 was not present in pulse synthesis. Therefore, if the polymerase transcribed the DNA faithfully it must have begun transcription within T_1GE 8. When oligonucleotide T_1GE 8 is present in its entirety it must have come from transcripts made by polymerase molecules that initiated at other points or traversed the entire SV40 circle.

When Ad2+ND3 DNA was transcribed by E. coli RNA polymerase and the transcript hybridized to SV40 DNA, the T_1 RNase fingerprints obtained corresponded to the site of initiation of E. coli RNA polymerase and the sequence beyond it (5), while the nucleotides derived from the sequence shown in the present report were absent, or present in relatively small amounts (Fig. 4). This is most easily understood if the E. coli RNA polymerase started at the same place on the SV40 DNA in Ad2+ND3 as it did on intact SV40 DNA.

DISCUSSION

The derivation of the sequence of the T_1 RNase digestion products was straight forward. The partial digestion products obtained with "C" RNase, carboxymethalyated pancreatic RNase, and spleen acid RNase were sufficient to obtain the complete order of the T_1 RNase digestion products. This sequence provided an independent confirmation of the "Late" (L) strand sequence presented in the preceding note. The sequences appear to contain the promotor sequences for E. coli RNA polymerase since the polymerase starts at the same place in the SV40 DNA segment of Ad2+ND3 as it does on intact SV40 DNA. After the present work was completed we learned of the work of Allet et al (12) indicating that the cleavage site for Hemophilus influenzae (Hin) restriction endonucleases (13) between SV40 DNA fragment Hin B and Hin G (14) may be part of a binding site for E. coli RNA polymerase on SV40 DNA. The cleavage site of Hin endonuclease is in the center of the hexanucleotide G-U-U-A-A-Cp (15). An interesting feature of this region of the sequence is a palindrome C-A-A-U-U-G-U-U-G-U-U-G-

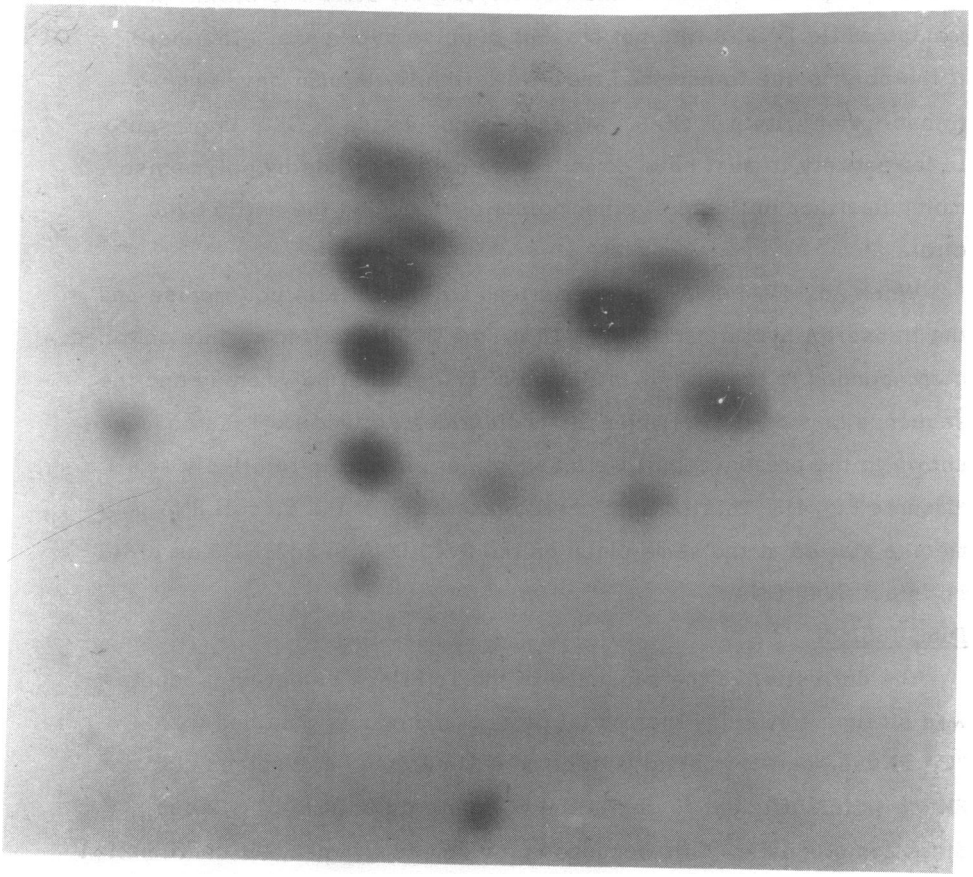
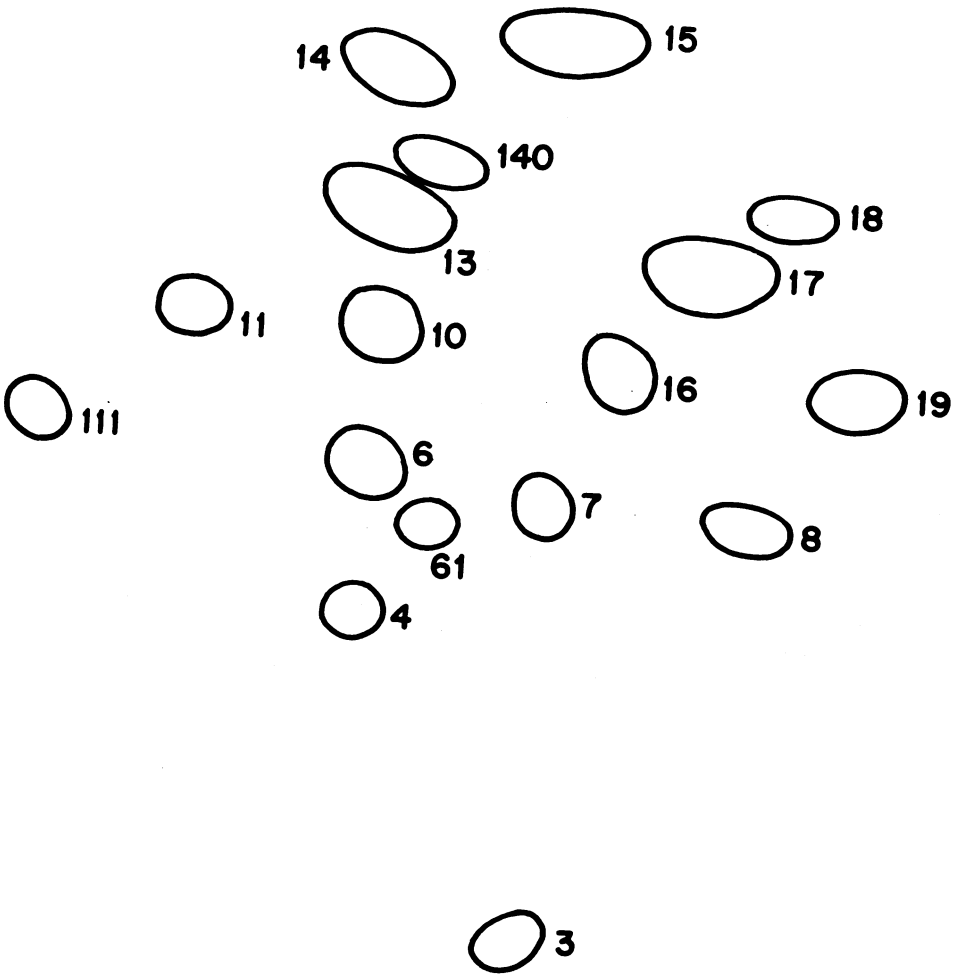


Figure 2

A. Autoradiograph of two dimensional electrophoretic separation of oligonucleotides produced by pancreatic RNase digestion of SV40 complementary RNA annealed to Ad2+ND3. Electrophoretic conditions were the same as in Fig 1 A and the RNA was the same preparation used in the experiment of Figure 1.



B. Schematic diagram of autoradiograph shown in Fig. 2A

TABLE 2

PRODUCTS OF U2 RNASE DIGESTION OF OLIGONUCLEOTIDES PURIFIED FROM EXHAUSTIVE T1 RNASE DIGESTS OF SV40 cRNA HYBRIDIZED TO Ad2+ND3

Oligonucleotide	α-^{32}P labelled precursor	U2 RNase Products R_b (a)	Sequence of product	Sequence deduced for initial oligonucleotide
T_1GE 114	GTP	0.17	CUUG	UUAACUUG
	ATP	0.42	UUA	
		0.15	UUAA	
T_1GE 553	GTP	0.37	UUG	UUUAUUG
T_1GE 8	ATP	0.35	UUA	UUACAAAUAAAG
		1.09	CA	
		0.95	UA	
		0.27	UAA	

(a) R_b refers to the electrophoretic mobility of the product relative to the dye xylene cyanol on DEAE paper at pH 3.5.

Table 2

The RNA sample was digested with T_1RNase and resulting oligonucleotides fractionated as in Figure 1. The products were further digested with U_2 RNase. The digestion products were analyzed on DEAE cellulose paper by electrophoresis at pH 3.5, and the resulting products were further analyzed by digestion with pancreatic RNase and electrophoresed as in Table 1.

U-U-A-A-Cp, (nucleotide No. 18 to 34, Fig. 3) with eight nucleotides on each side and a pivot Gp in the center. The transcript of the Hin restriction endonuclease cleavage site is contained at one end of this palindrome and a hexanucleotide C-A-A-U-U-Gp that corresponds to a symmetric hexanucleotide region in the DNA lies at the other end. Since the polymerase transcribes in only one direction there must be an assymetric feature to the initiation region. The symmetric site provided by a restriction endonuclease recognition sequence would not in itself appear sufficient to specify strand selection. Among many possibilities, the proximity of two different symmetrical sites in the DNA, as occurs in the present sequences could provide both binding sites and directional information for the enzyme.

Because of the large size and relatively rapid processing of nuclear SV40 cRNA, information about initiation sites for transcription of viral DNA in infected cells is lacking, though the 5' terminus of cytoplasmic "E" strand SV40 cRNA lies near the Hin A Hin C junction, far removed from the RNA polymerase initiation site whose sequence is reported here. Also at least two other (α-^{32}P) ATP labelled initiation sequences were detected in transcripts of SV40 DNA prepared with E. coli RNA polymerase (Lebowitz, P. personal communication). In a homologous system, there are at least three initiation sites in vitro for E. coli polymerase on fd RF-I DNA(16), a double stranded circular DNA of size similar to SV40 DNA. Additional promotors that require positive control elements such as 3'5' cyclic Ap could not be detected in any of these experiments. Therefore it is probably premature to draw conclusions about the possible function of the DNA sequence derived here in initiating transcription in infected cells.

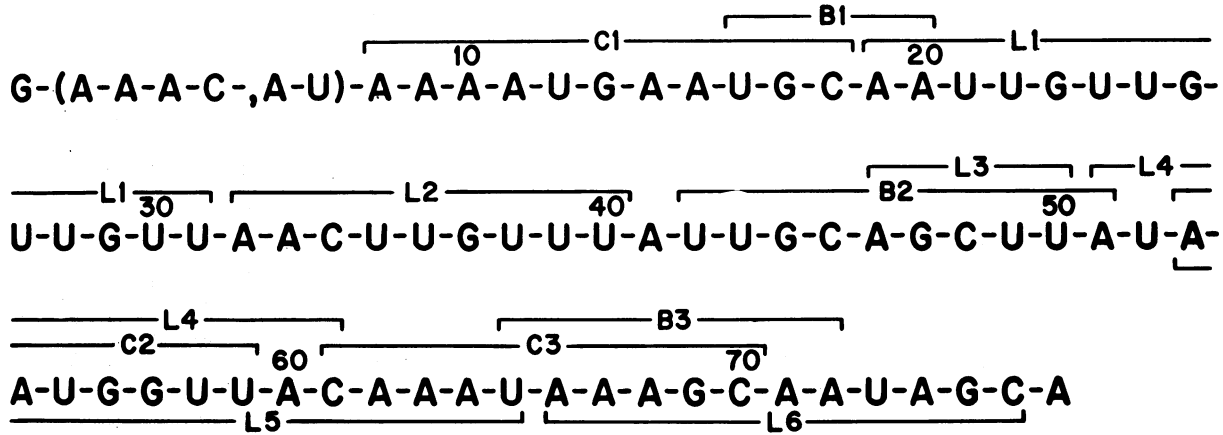


Figure 3

Nucleotide sequence of the transcript of "Early" strand of SV40 DNA preceding the E. coli RNA polymerase start within Ad2+ND3 DNA.

TABLE 3

PRODUCTS OF T1 RNASE TREATMENT OF OLIGONUCLEOTIDES FROM SV40 cRNA WHICH HAD BEEN HYBRIDIZED TO Ad2+ND3 DNA ELUTED AND DIGESTED WITH PANCREATIC RNASE

Oligonucleotide	T1 RNase digestion products				Sequence deduced
	α-^{32}P ATP	α-^{32}P CTP	α-^{32}P GTP	α-^{32}P UTP	
AGE 1	AAAC	AAAC	-	-	AAAC(A)
AGE 13	AU	-	-	AU	AU(A) AU(A)
AGE 4	AAAAU	-	AAAAU	AAAAU	AAAAU(G)
AGE 7	G, AAU	-	AAU	AAU	GAAU(G)
AGE 15					GC(A)
AGE 10	AAU		AAU	AAU	AAU(U) AAU(G) AAU(A)
AGE 18	-	-	-	G, U	GU(U)
AGE 56	AAC	AAC			AAC(U)
AGE 11	-	AG	AG	C	AGC(U)
AGE 19	-	-	G	G, U	GGU(U)
AGE 23	-	AC	-	AC	AC(A)
AGE 6	AAAU	-	-	AAAU	AAAU(A)
AGE 24	C, AAAG	AAAG	AAAG	-	AAAGC(A)

Table 3

The RNA samples were digested with pancreatic RNase and the resulting products obtained as mentioned in Table 1 were further digested with T_1RNase, and the resulting products identified by their electrophoretic mobility on DEAE paper at pH 3.5 and 1.7. Where necessary, products were also analyzed by alkaline hydrolysis.

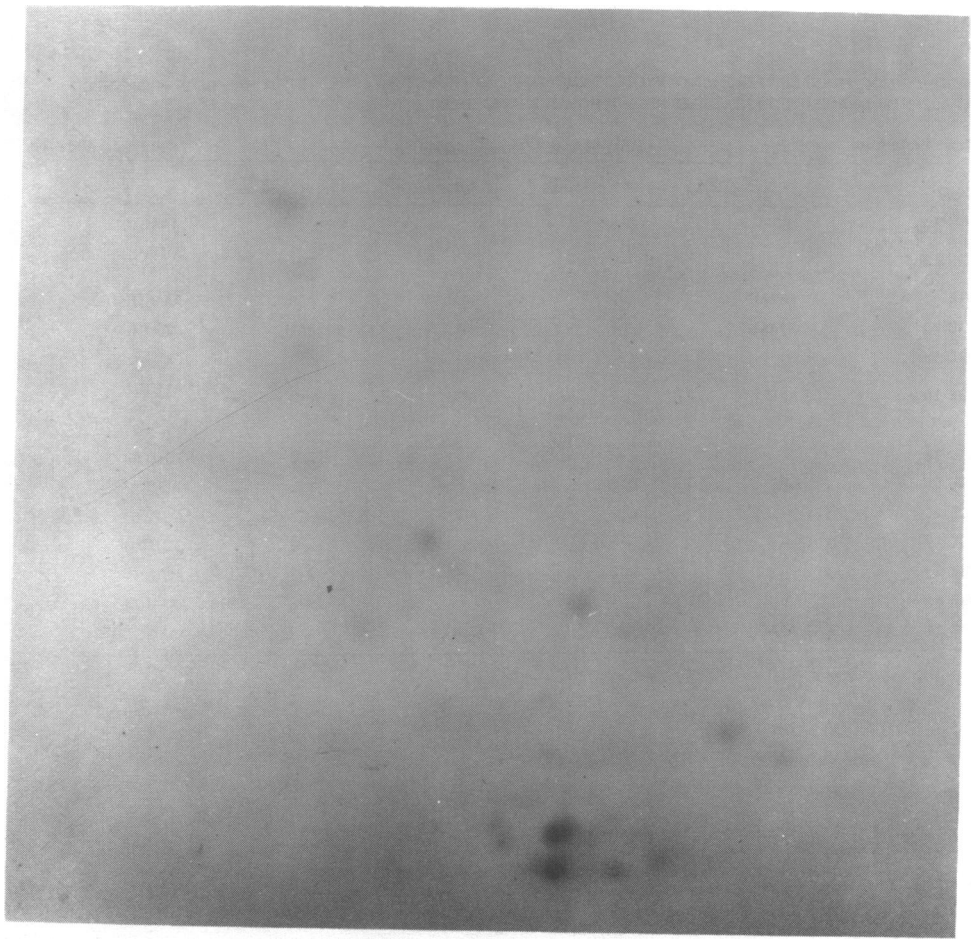
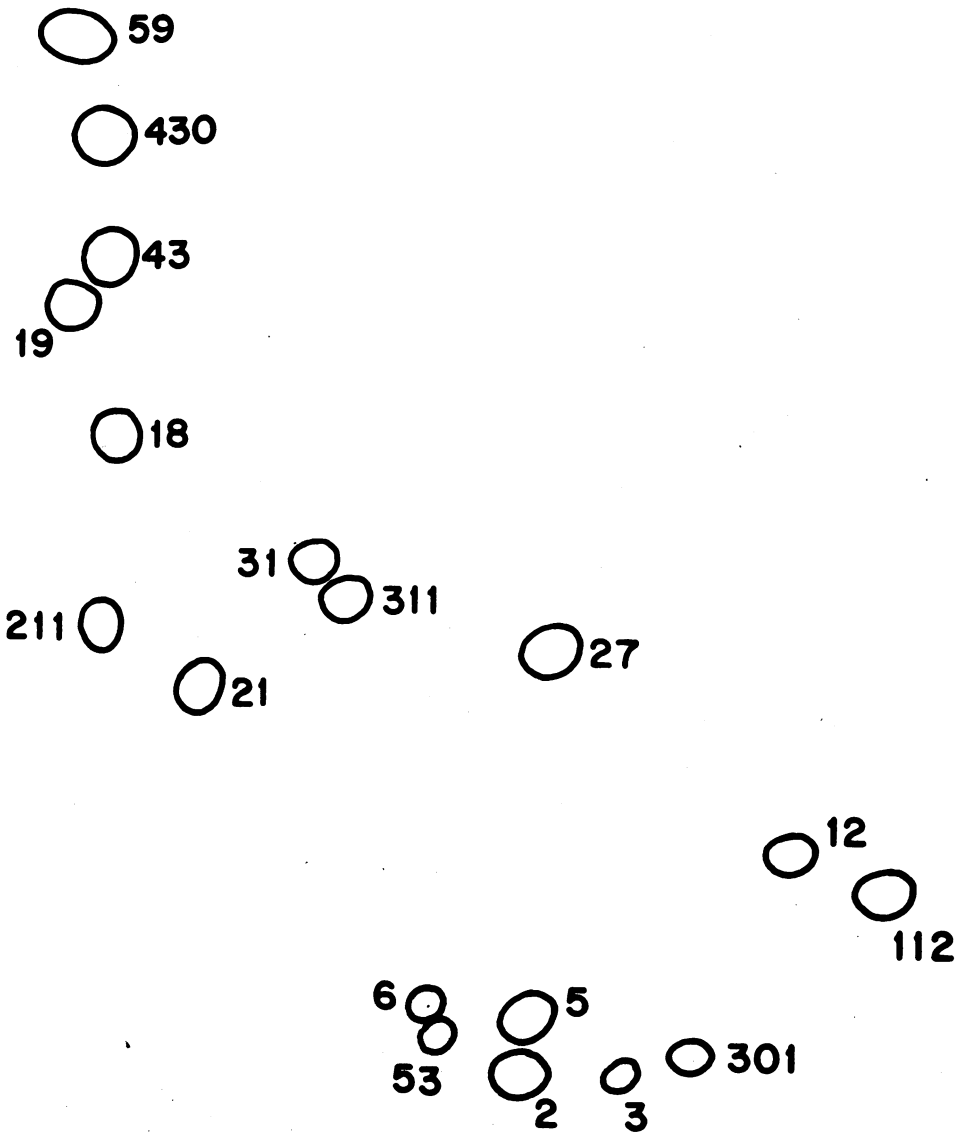


Figure 4

A. Autoradiograph of RNA prepared and analyzed as in Fig. 1A, except that RNA was synthesized with Ad2+ND3 DNA as template and the transcript annealed to SV40 DNA, RNA was synthesized in presence of α-^{32}P ATP (30 Ci/m Mole).



B. Schematic sketch of autoradiograph of Fig. A

TABLE 4

PRODUCTS OF LIMITED DIGESTION OF C SV40 RNA HYBRIDIZED TO Ad+ND3 ELUTED, AND TREATED WITH "C" RIBONUCLEASE

Partial digestion products	Products from extensive T1 RNase digestion		Sequence deduced
	R_b 1.7	Sequence of Product	
C_1	0.22	AAAAUG (A)	AAAAUGAAUGC(A)
	0.42	AAUG (C)	
	2.1	C(A)	
C_2	0.52	AAUG (G)	AAUGGUU(A)
	0.79	UU(A),	
	1.8	G(U)	
C_3	0.10	AAAUAAAG(C)	AAAUAAAGC(A)
	2.1	C(A)	

Table 4

Separate samples of RNA were prepared with each of the four α-^{32}P nucleoside triphosphates. The RNA was digested with "C" ribonuclease (7) and the resulting oligonucleotides were separated by electrophoresis on cellogel strips at pH 3.5 and homochromatography on DEAE cellulose thin layer plates with "Homochromo B" or "A" (17). The individual products were eluted and digested with T_1 RNase and the digest was analyzed by electrophoresis on DEAE cellulose paper pH 1.7 and the resulting oligonucleotides were further digested with pancreatic RNase and analyzed by electrophoresis on DEAE cellulose paper at pH 3.5. Results with all four triphosphates were in agreement with the sequences shown. Brackets indicate the nearest neighbor nucleotide. R_b 1.7 refers to the electrophoretic mobility of a product relative to the dye xylene cyanol on DEAE paper at pH 1.7.

TABLE 5

PRODUCTS OF LIMITED DIGESTION OF SV40 cRNA HYBRIDIZED TO Ad2+ND3 DNA ELUTED, AND TREATED WITH CARBOXYMETHALATED PANCREATIC RNASE

Partial digestion products	Products of extensive T1 RNase digestion		Sequence deduced
	R_b 1.7	Sequence of Product	
L_1	0.209 0.38 0.78	AAUUG(U) UUG (U) UU(A)	AAUUGUUGUUGUU(A)
L_2	0.22 0.14	AACUUG(U) UUU(A)	AACUUGUUU(A)
L_3	1.6 0.6	AG(C) CUU(A)	AGCUU(A)
L_4	0.00 0.72	AUAAUG(G) UUAC(A), G(U)	AUAAUGGUUAC(A)
L_5	0.4 0.15	AAUG (G) UUACAAAU(A) (G)	AAUGGUUACAAAU(A)
L_6	1.00 0.36	AAAG (C) CAAUAG (C) C(A)	AAAGCAAUAGC(A)

Table 5

Radioactive RNA was prepared as described in Table 4. The RNA was digested with carboxymethylated pancreatic RNase (9, 10) and further analysis of the resulting oligonucleotides was done as in Table 4. Lines underscoring a product indicate molar yield in excess of one.

TABLE 6

PRODUCTS OF LIMITED DIGESTION OF SV40 cRNA HYBRIDIZED TO Ad2+ND3 ELUTED AND DIGESTED WITH SPLEEN ACID RIBONUCLEASE

Partial digestion products	Products from extensive digestion with T1 RNase		Sequence deduced
	R_b 1.7	Sequence of Product	
B_1	0.85 1.181	UG(C) CAA(U)	UGCAA(U)
B_2	0.37 1.29 0.7	UUG(C) CAG (C) (CUUA)	UUGCAGC(UU(A))
B_3	0.44 1.94	UAAAG (C) CAA(U)	UAAAGCAA(U)

Table 6

Radioactive RNA was prepared as described in Table 4. The RNA was digested with spleen acid (RNase (11) and the resulting oligonucleotides analyzed as in Table 4.

ACKNOWLEDGEMENTS

We thank Su Fang Su, Francis DeNoto, Lucy Price and Benjamin Jones for excellent technical assistance. This work was supported by Grant Number CA 10137-08 from the National Institute of Health and VC 1E from the American Cancer Society.

REFERENCES

1 Westphal, H. (1970) J. Mol. Biol. 50, 407-420
2 Zain, B.S., Dhar, R., Weissman, S.M., Lebowitz, P., Lewis, A.M. Jr. (1973) J. Virol. 11, 682-693
3 Lewis, A.M. Jr., Levine, A.S., Crumpacker, C.S., Levine, M.J., Samaha, R.J., and Henry, P.H. (1973) J. Virol. 11, 655-664
4 Zain, B.S., Weissman, S.M., Dhar, R., Pan, J. submitted for publication
5 Dhar, R., Zain, B.S., Weissman, S.M., and Pan, J., (1974) Proc. Nat. Acad. Sci., (in press)
6 Arima, T., Uchida, T., and Egami, F. (1968) Biochem. J., 106-601
7 Levy, C.C., and Goldman, P. (1970) J. Biol. Chem. 245, 3257-3262
8 Marotta, C.A., Levy, C.C., Weissman, S.M., and Varrichio, F. (1973) Biochemistry, 12, 2901-2904
9 Contreras, R., and Fiers, W. (1971) Fed. Eur. Biochem. Soc. Letters, 6, 281-283
10 Goldstein, J. (1967) J. Mol. Biol., 25, 123-130
11 Bernardi, G., in G.L. Cantoni and F.R. Davies (Editors) (1966) Procedures in Nucleic Acid Research p. 37, Harper and Row, New York
12 Allet, B., Roberts, R.J., Gesteland, R.F., Solem, R. (submitted for publication)
13 Smith, H.O., and Wilcox, K.W. (1970) J. Mol. Biol. 51, 379-392
14 Danna, K., Sack, G.M., and Nathans, D. (1973) J. Mol. Biol. 78, 363-376
15 Kelly, T.J., and Smith, H.O., (1970) J. Mol. Biol., 51, 393-409
16 Takonomi, M. Okamoto, T., Sugiuro, M. (1970) Cold Spring Harbor Symp. Quant. Biol. 35, 179-187
17 Brownlee, C.G., and Sanger, R. (1969) Eur. J. Biochem., 11, 395-408

* Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA