Volume 1 number 4 April 1974 Nucleic Acids Research

The nucleotide sequence preceding an RNA polymerase initiation site on SV40 DNA. Part 1. The sequence of the late strand transcript

B.S.Zain, S.M.Weissman, R.Dhar, and J.Pan

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 transcript of the "late" strand of the region of SV40 DNA preceding the preferred initiation site for Escherichia coli RNA polymerase has been determined to be U-G-U-A-A-C-C-A-U-U-A-U-A-A-G-C-U-G-C-A-A-U-A-A-A-C-A-A-G-U-U-A-A-C-A-A-C-A-A-C-A-A-U-U-G-Cp. Hemophilus influenza restriction endonuclease cleaves this region 30 nucleotides (base pairs) before the site of initiation of RNA synthesis by RNA polymerase.

INTRODUCTION

Knowledge about the regions of DNA involved in the initiation and termination of transcription by RNA polymerase is essential to understand the mechanism of regulation of gene expression. At present all that is known about possible start signals is that transcription by E. coli nucleotidyl transferase (E.C. 2.7.7.6 RNA polymerase) begins with a purine; while the second and subsequent bases are variable(1). Since several transcripts prepared with E. coli RNA polymerase end with U-U-U-U-U-U-Ap(2, 3), this may be part of a termination signal. The sequences U-U-U-U-U-U or U-U-U-U-U purine cannot be a sufficient signals for termination in all situations since such sequences occur internally in RNA transcribed from coliphage T_4 DNA(4) or SV40(8). There is almost no information about the nucleotide sequences preceding transcription initiation sites or following recognized transcription termination sites although these presumably play a critical role in specifying the sites.

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.

SV40 virus is an oncogenic DNA virus with a genome of approximately 3.5×10^6 molecular weight. Because of its relative simplicity it has become an object of extensive study by several laboratories using both biochemical and genetic approaches. We have been involved in detailed sequence studies of SV40 transcripts prepared in vitro and in infected cells. E. coli RNA polymerase has been known for some time to transcribe Form 1 SV40 DNA assymetrically, transcribing only the "early" (E) strand (5). In an earlier communication we reported the presence of a preferred initiation site for E. coli RNA polymerase located within the SV40 segment of the Adeno-2 SV40 hybrid viruses, Ad2+ND1 and Ad2+ ND3(6). The initiation site is also contained in the G fragment (SV40 Hin G) produced by cleavage of SV40 DNA with Hemophilus influenza restriction enzymes (6, 7). In contrast to transcription of SV40 Form 1 DNA, the H. influenza restriction endonuclease generated SV40 Hin G fragment is transcribed selectively from the "L" strand (8). This "L" strand transcript overlaps the "E" strand initiation site for E. coli RNA polymerase. We have used this transcript to derive the nucleotide sequence preceding the polymerase start. Independently we derived the sequence of the E strand transcript from this region and confirmed that the two sequences were exactly complementary(9).

MATERIALS AND METHODS

Most of the materials needed for sequence analysis were obtained from commercial sources as described previously (2, 6). U2 ribonuclease (E. C. 2. 7. 7. 26), "C" enzyme (10), spleen acid ribonuclease (11) were the generous gifts of Sankyo Co., Dr. G. Bernardi, Dr. K. Levy and Dr. P. Lebowitz. H. influenza restriction endonucleases were prepared by the method of Smith and Wilcox (12) and E. coli RII restriction enzyme by the method of Yoshimori and Boyer (13). Details of cell and virus propagation, DNA extraction, transcription of DNA (using E. coli RNA polymerase either at 37°C or 18-24°C after synchronization of

the initiation of RNA synthesis) preparation of E. coli RNA polymerase, nucleic acid hybridization and sequence analysis have been described in detail elsewhere (2, 6, 14). SV40 DNA was prepared by the Hirt extraction procedure from cells infected with 777 virus obtained from Dr. M. Oxman, or from a small plaque forming virus provided by Dr. D. Nathans. The cleavage products obtained by digestion of (^{32}P) labelled SV40 DNA with the restriction enzymes were resolved by electrophoresis on 4% acrylamide slab gels (40 cm x 0.3 cm) as described earlier (7, 14). The DNA was extracted from the gel by homogenization in 0.15 M NaCl 0.015 M Na Citrate (pH 6.7), and concentrated by alcohol precipitation. To prepare radioactive SV40 cRNA, 2-3μg of RNA polymerase holoenzyme was incubated with template DNA (0.1-0.2μg of Hin fragment DNA or 1-3μg of intact supercoiled SV40 DNA) in a reaction mixture containing 0.09 M KCl, 0.033 M Tris/HCl (pH 7.9), 6mM 2- mercaptoethanol, 3.3 mM $MgCl_2$, 0.165 mM XTP in a total volume of 64μl. The reaction was allowed to proceed at 37°C for 30 minutes, followed by addition of 0.2μg of pancreatic DNase (ribonuclease free-Worthington) and incubation at 37°C for 10 minutes. For limited synthesis, the reaction mixture was constituted as for standard synthesis, but with the omission of $MgCl_2$ and reduction of KCl to 2.5 mm. After preincubation for 1-2 minutes at 25°C, $MgCl_2$ was added to a concentration of 6.1 mM and the reaction was allowed to proceed for 5-60 seconds at 18-24°C. After incubation, the reaction was stopped by the addition of SDS to 0.25% and an equal volume of water saturated redistilled phenol. The RNA was extracted and passed through a Sephadex G-100 column, to remove the unincorporated nucleoside triphosphates. The eluted RNA was precipitated in the presence of 50-100μg of nonradioactive tRNA carrier by the addition of 2 volumes of alcohol. The RNA was either directly analyzed or annealed to DNA immobilized on nitrocellulose filters (18 hrs., 67°C 0.3 M NaCl, 0.03 M Na citrate, pH 6.7, total volume 2 ml) eluted with boiling water for 5 min., and reprecipitated before analysis.

For sequence analysis, RNA labelled with a single radioactive triphosphate was digested with T_1 (E.C. 3.1.4.8) or pancreatic RNase (1-A, E-C 2.7.7.16). The products were fractionated by electrophoresis on Cellogel (pH 3.5) and homochromatography on DEAE cellulose TLC plates with a 3% RNA solution. The radioactive oligonucleotides were digested with either T_1, pancreatic or U2 RNAse, and the products further analyzed by electrophoresis on DEAE paper at pH 3.5 or 1.7 and by alkaline hydrolysis. Ordering of the oligonucleotides was determined by partial digestion of the RNA with T_1 RNase, spleen acid RNase or "C" RNase. The resulting products were resolved by electrophoresis on Cellogel strips and homochromatography on DEAE cellulose plates in a 5% solution of partially hydrolyzed yeast RNA. The partial digestion products were analyzed by complete digestion with T_1 RNase or pancreatic RNase followed by electrophoresis on DEAE paper at pH 1.7. The resulting oligonucleotides were further digested with T_1 RNase, pancreatic RNase or U2 RNase and the digestion products were identified by electrophoresis on DEAE paper at pH 3.5, and by alkaline hydrolysis.

RESULTS

Transcription of the SV40-Hin G fragment in vitro followed by digestion of the transcript with either T_1 (Fig.1) or pancreatic RNase (Fig. 2) gave consistent sets of oligonucleotides. The products were different from those observed when transcripts from whole SV40 DNA were annealed to Hin G fragment immobilized on filters, suggesting that the Hin G transcript might be derived from the "L" strand of SV40. Subsequent sequence analysis of both sets of oligonucleotides has fully confirmed this expectation. Transcripts of the Hin G fragment derived from the "E" strand were present in only very small amounts, and generally contained less than 10% as much radioactivity per nucleotide as the transcripts from the "L" strand. The nucleotide sequences of nearly all the pancreatic ribonuclease products of the Hin G transcript could be determined directly by comparison of T_1 ribonuclease digests of RNA separately labelled with each of the

four (α-^{32}P) triphosphates. To derive the sequences of T_1 ribonuclease products it was in general necessary to resort to both pancreatic ribonuclease and limited U2 ribonuclease digestion of each product, in addition to nearest neighbor analysis.

It was possible to group the Hin G fragment sequences into two subsets by cleavage of the fragment with E. coli RII restriction endonuclease. This nuclease cleaved Hin G into products Hin G RII-1 and Hin G RII-2. The latter product has been shown by redigestion (15), hybridization and sequence analysis to be a portion of the RII F fragment produced by cleavage of SV40 DNA. Since RII F overlaps Hin B, the DNA fragment Hin G RII-2 must represent that portion of Hin G closest to the Hin B fragment. Hin G RII-2 was found to contain the nucleotide sequence from the preferred initiation site for E. coli polymerase on SV40 DNA to position approximately 116 from the triphosphate end of limited transcripts (8). Sequences beyond position 116 were missing, indicating that Hin G had been cleaved at approximately this site. This corresponds to the deduction of others (16, 17) that the RII restriction cleavage site has the sequence C-C-A-G-Gp.

In addition to late strand products complementary to the early part of the pulse labelled SV40 transcript, certain sequences not complementary to pulse labelled SV40 RNA were detected in the transcripts of Hin G and Hin G RII-2. Those from the latter fragment contained oligonucleotides T_1GL 9, 12, 22, and 44 (Figure 1, Table 1). These proved to be the sequences extending from the point of cleavage of SV40 DNA by the Hin restriction endonuclease to the initiation site of E. coli RNA polymerase. Data presented in Tables 1, 2, and 3 established the sequences of the T_1 RNase products and the pancreatic RNase digestion products from this region. When the transcript of Hin G fragment was digested with spleen acid ribonuclease, a prominent product was an oligonucleotide (B1)

with the sequence U-A-A-G-C-U-G-C-A-A-(U)p that could only have arisen from an overlap between the T_1GL products 9, 44, and 12. This oligonucleotide was isolated separately from digests individually labelled with each of the (α-^{32}P) triphosphates and analyzed by digestion with T1 and pancreatic ribonuclease (Table 4). Products were identified by their electrophoretic mobility and by further redigestion with pancreatic or T_1 ribonuclease respectively. This was sufficient to determine much of the sequence shown in Figure 4. In addition, products obtained by cleavage of the G fragment transcript with "C" enzyme included an oligonucleotide with the sequence A-A-U-G-U-U-U-(A)p. The only source for this product could have been the 5' end of the sequences shown in Figure 4. Also, the complement of the 5' end of the sequence in Figure 4 G-U-U-A-Cp could have been derived from only one of the T_1 products obtained from SV40 transcript which had been annealed to Hin G. This product had the sequence U-U-A-C-A-A-A-U-A-A-A-G-Cp (see part 2) and contained within it the sequence A-A-A-U-A-A-A-Gp that had been found to be the nucleotide sequence of the gamma P^{32} labelled T_1 product obtained from pulse synthesis of SV40 DNA annealed to Adeno 2+ ND1 or Adeno 2+ ND3. While there was another oligonucleotide with the same 3' sequence found in E strand transcripts of the Hin G fragment this latter T_1 oligonucleotide was shown to be present internally in limited transcripts of SV40 (8) so that it could not have contained the initiation site. Therefore, the 5' end of the sequence of the late strand transcript shown in Figure 3 overlapped the 5' end of the early strand transcripts as seen in pulse synthesis from SV40 DNA and the sequence of Figure 4 provided the entire sequence for the nucleotides from the H influenzae restriction cut in SV40 DNA to the site of E. coli polymerase start.

To obtain nucleotide sequences from the B fragment L strand transcripts which immediately precedes the junction of the G and B fragments we transcribed separated strands of the Adenovirus SV40 hybrid virus Ad 2+ ND3A DNA (18) kindly provided by Dr. S. Patch,

A. Levine, and A. Lewis. A transcript of the Ad 2+ ND3A strand which contained the L strand of a segment of SV40 DNA was then annealed to SV40 DNA which was immobilized on a filter. Oligonucleotide maps of T_1 RNase digests of the resulting RNA (Fig. 3) showed one principal product (T_1 BL-7) in addition to those derived from the Hin G fragment. This product was found to have the sequence U-U-A-A-C-A-A-C-A-A-C-A-A-U-U-Gp. The 5' end of this fragment would be G-U-U-A-A-Cp, one of the sequences previously shown to be an H influenzae cleavage site (19). More direct evidence joining this oligonucleotide to the sequence derived from the G fragment transcript was obtained by comparing E strand oligonucleotide sequences from SV40 transcript. One and only one oligonucleotide was obtained which could be complementary to the L strand sequences described immediately above. This oligonucleotide had the sequence G-U-U-A-A-C-U-U-G-Up, hence provided the overlap of oligonucleotide T_1GL-12 and the oligonucleotide just described. Therefore, the sequence shown in Figure 4 represents the transcript of the L strand of SV40 DNA extending from the initiation site for E. coli RNA polymerase through the junction between the Hin G and Hin B fragments of SV40 DNA.

DISCUSSION

In the course of this work we transcribed and analyzed DNA from several sources that overlapped the sequences described here. These included DNA from the separated strands of Ad2+ND3 virus DNA from the Hin G fragment of SV40 DNA, from the ZC fragment, produced by digestion of SV40 DNA with H. aegyptius restriction enzymes (15, 20, 21, 18) (see below) and DNA from the E. coli RII-F fragment of SV40. We further compared sequences obtained from RNA synthesized in infected cells with that obtained from the in vitro transcripts of SV40. In each case the sequence in this region proved to be identical. Further, the same sequence was found whether the SV40 virus used was the mixed stock 777 virus or whether it was derived by low multiplicity passage from a small

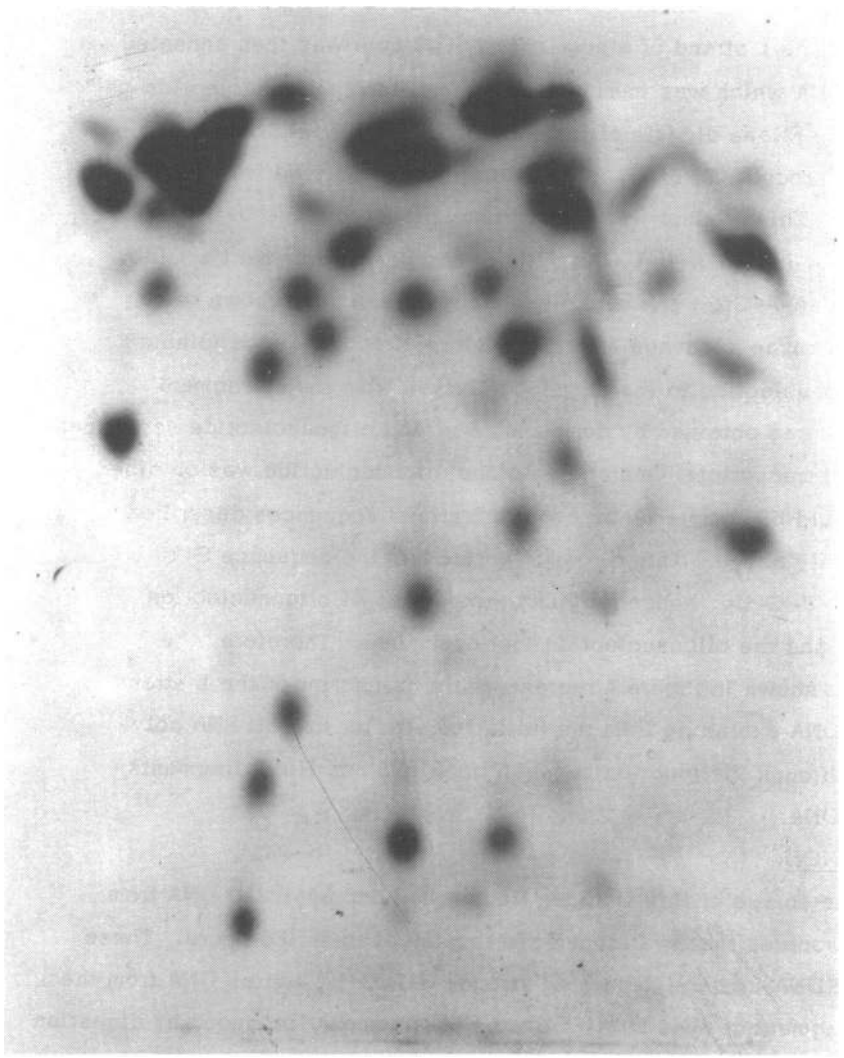

Figure 1
Autoradiograph of a two dimensional electrophoretic separation of oligonucleotides produced by T_1 RNase digestion of Hin G cRNA. Electrophoresis was from left to right on cellogel in 7 M urea at pH 3.5 and from bottom upwards on a DEAE cellulose (9:1) thin layer chromatography plate with the "Homo B" solution of Sanger et al (27). The RNA was synthesized with SV40-Hin G fragment DNA template, in the presence of (α-^{32}P)GTP (125 Ci/m mol) at 37 °C for 30' as described in the Methods section.
(a) Autoradiograph
(b) Schematic sketch of autoradiograph with numbers assigned to only those oligonucleotides from which the sequence shown in Fig.4 was derived.

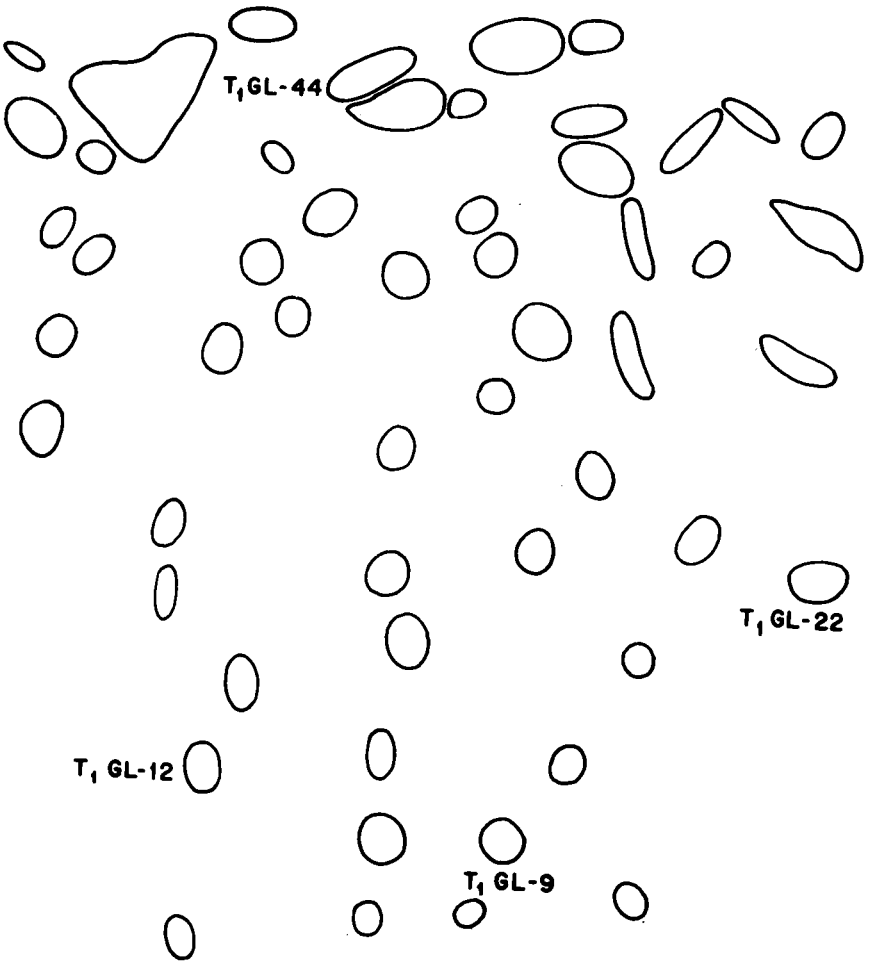

TABLE 1

ANALYSIS OF OLIGONUCLEOTIDES OBTAINED BY COMPLETE DIGESTION OF cRNA WITH T1 RNase

Oligonucleotide Number	Source of RNA	Pancreatic RNase digestion Products [a] Radioactive Precursors Used to Label cRNA α^{32}P-ATP	α^{32}P-CTP	α^{32}PGTP	α^{32}PUTP	Sequence Derived
T_1GL-12	Hin G	AAAC, AAG, AAU, C	AAAC	AAG	AAU, AAG	C(AAUAAAC)AAG(U)
T_1GL-44	Hin G	-	G	U	C	CUG(C)
T_1GL-9	Hin G	AAG, AAC, AU U, C	AAG, AAC	AAG	AU	U(AACC, AUU, AU)AA G(C)
T_1GL-22	Hin G	U	-	U	AU, C, G, U	C(U, U, U, AU, U) UG(U,)
T_1BL-7	Ad2+ND3 "L" strand	AAC, AAU, U	AAC, G	U	AAU, U	(U, UAACAACAAC, AAU)UG(Cp)

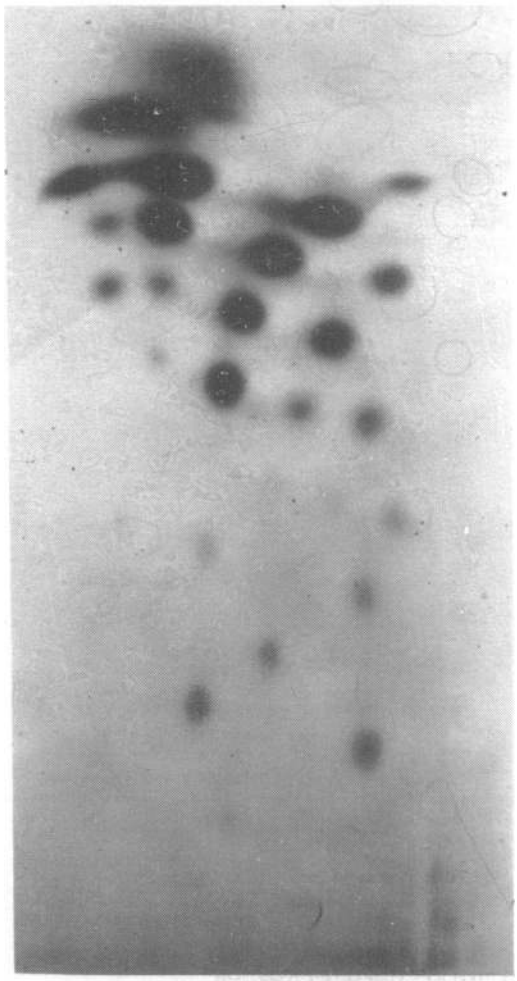

Figure 2

Autoradiograph of a two dimensional electrophoretic separation of oligonucleotides produced by pancreatic RNase digestion of Hin G cRNA, fractionated as described in Fig.1. The RNA synthesized was labelled with (α-^{32}P) UTP (98 Ci/m M) at 37 °C for 30' as described in Methods.

(a) Autoradiograph

(b) Schematic sketch of the autoradiograph with numbers assigned only to those oligonucleotides from which the sequence shown in Fig.4 was derived.

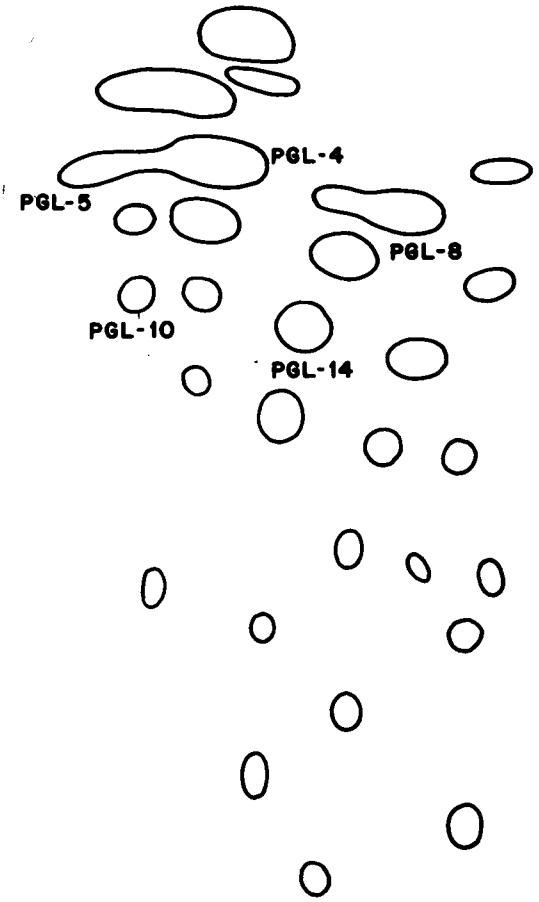

TABLE 2

ANALYSIS OF OLIGONUCLEOTIDES OBTAINED BY COMPLETE DIGESTION OF HIN G cRNA WITH PANCREATIC RNase

Oligonucleotide Number[b]	T_1 RNase digestion Products[a] Radioactive Precursors Used to Label cRNA				Sequence Derived
	^{32}P ATP	^{32}P CTP	^{32}P GTP	^{32}P UTP	
PGL-14	-	-	AAG	AAG	AAGU-OH
PGL-α	AAC	AAC	-	-	AAC(A)
PGL-3	C	G	-	-	GC(A)
PGL-10	AAG	AAG	AAG	C	AAGC(U)
PGL-4	-	-	-	AU	AU(U), AU(Ap)
PGL-3C	AAC	AAC	-	-	AAC(C)
PGL-8	U	-	-	G	GU(A)
PGL-5	-	G	-	C	GC(U)

(a) The RNA was digested with T1 RNase or Pancreatic RNase. The resulting oligonucleotides were separated by electrophoresis on cellogel at pH 3.5 and homochromatography on thin layer chromatography plates as described in Materials and Methods. Individual products were located by autoradiography, eluted and digested with Pancreatic RNase or T1 RNase. These digestion products were fractionated on DEAE paper by electrophoresis at pH 3.5 and the products identified by their electrophoretic mobilities, and, where necessary, by base analysis.
Lines underscoring a product indicate the nearest integral yield in excess of one. Brackets indicate the nearest neighbor base. (b) Oligonucleotide numbers in Tables 1 and 2 refer to Figures 1 and 2 respectively.

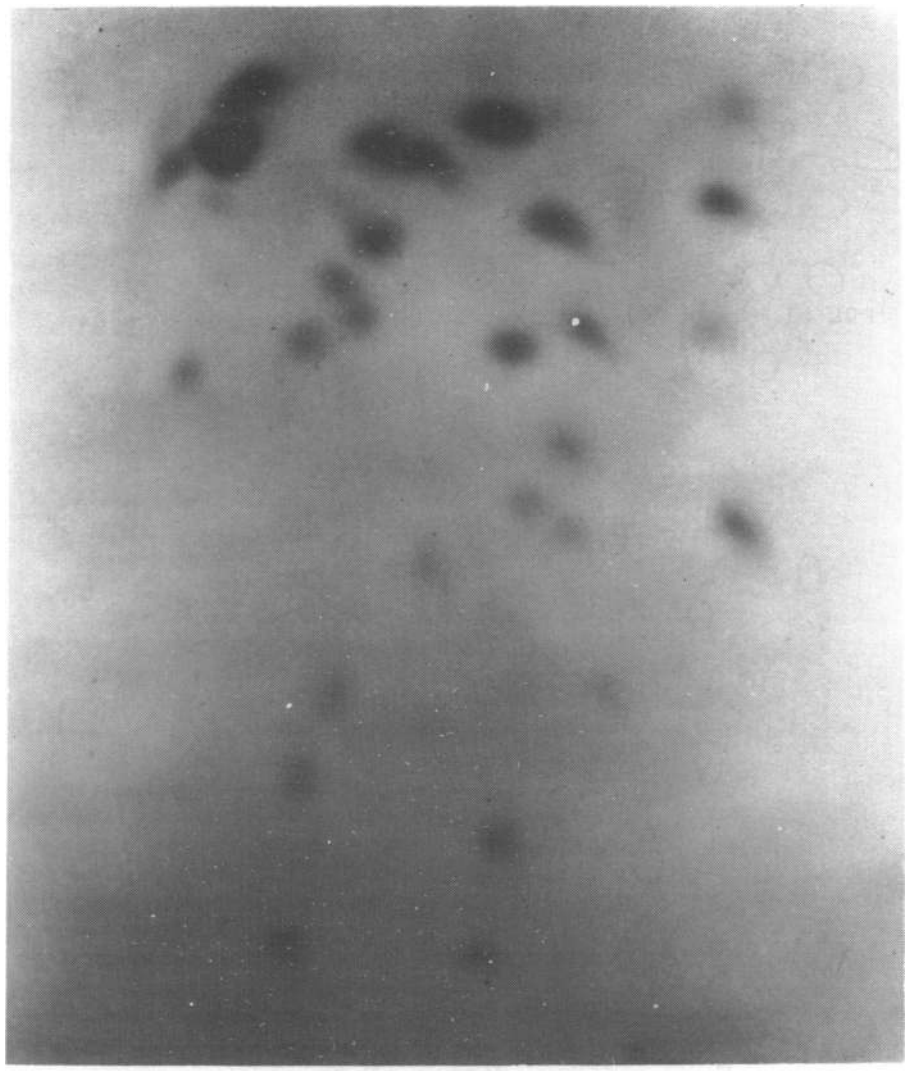

Figure 3

Autoradiograph of a two dimensional electrophoretic and chromatographic separation of oligonucleotides produced by T_1 RNase digestion of Ad2 + ND3A "L" strand cRNA, as described in Fig.1. The DNA template used for the preparation of RNA was "L" strand of Ad2 + ND3 DNA.

(A) Autoradiograph

(B) Schematic sketch of the autoradiograph with numbers assigned to those oligonucleotides from which the sequence shown in Fig.4 is derived.

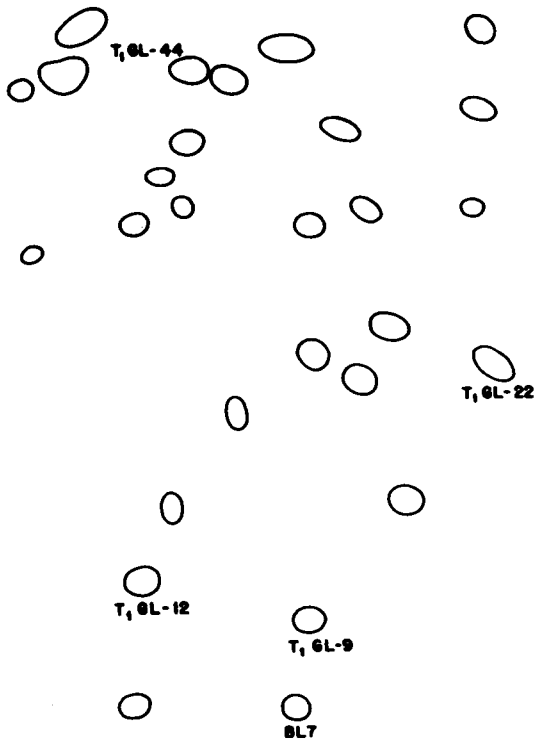

TABLE 3

ANALYSIS BY U2 RIBONUCLEASE DIGESTION OF VARIOUS OLIGONUCLEOTIDES OBTAINED BY DIGESTION OF HIN G cRNA WITH T1 RIBONUCLEASE

Oligonucleotide No. (a)	Labelled Precursor	RB (3.5)(b)	Product Identification (c)	Sequence Derived
T_1GL-22	(α-^{32}P)ATP	1.4/20	CUUUA	CUUUAUUUG(U)
T_1GL-9	(α-^{32}P)ATP	4.8/15 6/15 10/15 15/15	UAA UUA CCA UA	UAACCAUUAUAAG(C)
T_1GL-12	(α-^{32}P)ATP	17/15 15/15 6.5/15	CA(A) UA(A) CAA(G)	CAAUAAACAAG(C)
T_1BL-7	(α-^{32}P)GTP	7/22	UUG	UUAACAACAACAAUUG (C)

(a) The first three numbers refer to Figure 1, T_1BL-7 refers to Figure 3.
(b) R_b 3.5 refers to the electrophoretic mobility of the product at pH 3.5 on DEAE paper relative to xylene cyanol.
(c) In addition to their electrophoretic mobility at pH 3.5, the products were further analyzed by electrophoresis on DEAE paper at pH 1.7 and by digestion with pancreatic RNase.

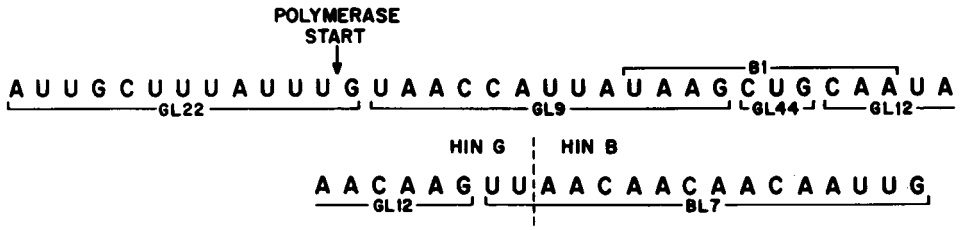

TABLE 4

ANALYSIS OF THE OLIGONUCLEOTIDE UAAGCUGCAA(U) PRODUCED BY LIMITED DIGESTION OF HIN G cRNA WITH SPLEEN ACID RIBONUCLEASE

Radioactive Precursor	RNase used for extensive digestion	Oligonucleotides in extensive digest	Analysis of Oligonucleotides from extensive digest			
			R_b (1.7)	Type of Analysis	Product of Analysis	Sequence of Oligonucleotides
(α-^{32}P)ATP	T_1	(a)	10.2/19	P 3.5	U, AAG	(UAAG)
		(b)	34/16.5	3.5	CAA	(CAA)
	Pancreatic	(a)	19.2/15.8	T_1 3.5	AAGC	(AAGC)
		(b)	27.5/15.8	3.5	GC(A)	GC(A)
		(c)	32/15.8	3.5	AA	AA
(α-^{32}P)CTP	T_1	(a)	3/14.5	P 3.5	AAG	UAAG(C)
		(b)	11/14.5	P 3.5	G	CUG(C)
	Pancreatic	(a)	20.5/16.7	T_1 3.5	AAG	AAGC
		(b)	28/16.7	T_1 3.5	G	CUG(C)
(α-^{32}P)GTP	T_1	(a)	6.5/14.3	P 3.5	AAG	UAAG
		(b)	9/14.3	P 3.5	U	CUG
	Pancreatic	(a)	17.5/15	T_1 3.5	AAG	UAAG(C)
		(b)		T_1 3.5	U	CUG(C)
(α-^{32}P)UTP	T_1	(a)	11/14.5	P 3.5	G	CUG
		(b)	30/14.5	3.5	CAA	CAA(U)

Table 4

The radioactive Hin G cRNA was partially digested with spleen acid RNase, followed by resolution of the products by two dimensional electrophoresis and homochromatography as described in methods. The resulting aligonucleotides (partial products) were analyzed first by extensive digestion with T_1 and pancreatic RNase (separately) followed by fractionation of the complete digestion products by electrophoresis on DEAE cellulose paper at pH 1.7. The products so obtained were further analyzed by digestion of the extensive T_1 RNase digestion products with pancreatic enzyme (or without any further digestion) followed by electrophoresis on DEAE paper at pH 3.5 (P 3.5 = Pancreatic RNase digestion followed by electrophoresis at pH 3.5, 3.5= no digestion but resolution by electrophoresis at pH 3.5). Similarly the extensive pancreatic RNase digestion products were analyzed by electrophoresis on DEAE paper at pH 1.7 with or without prior T_1 RNase digestion. The products were identified by their R_b (relative mobilities with reference to the blue dye, (xylene cyanol). Derivation of the sequence is based on the data shown in Table 1, 2, 3 and 4, and from the position of the partial product on the autoradiograph (figure not shown).

plaque isolate provided by Dr. K. Danna and D. Nathans. We therefore believe that the sequence represents an accurate transcript of SV40 DNA and furthermore that sequences within this region of SV40 DNA are preserved during repeated passage of this virus in cells. The fact that the Hin G fragment transcribes preferentially from the L strand is of some curiousity since this results not only from the destruction of an E strand initiation site by cleavage with the Hin enzyme but possibly from the generation of a preferred site for the initiation of transcription of the L strand. This provided a fortuitous opportunity to derive a sequence preceding the point where F. coli polymerase prefers to initiate synthesis on whole SV40 DNA. In the accompanying report these results have been confirmed and supplemented by analysis of the sequences derived from this region of the E. strand of SV40 DNA, obtained by polymerase read-through of the initiation site on the DNA.

The promotor has been defined as a region of DNA which is "an indispensable initiator element"(22) and "a site which serves to initiate the transcription of an operon"(23). Blattner and Dahlberg(24) called the region of DNA which includes polymerase entry drift and start site the promotor region. Thus the site in SV40 DNA which immediately precedes the site of initiation of RNA synthesis and is not transcribed under conditions of limited synthesis would constitute part of the promotor region. Not all the promotor region is present in Hin G. However when we transcribed either the H. aegyptius restriction enzyme fragment ZC of SV40 DNA (a fragment which extends approximately 90 nucleotides to the 5' end of the L strand transcript of Hin G (15, 21), or Ad2+ND3 DNA (which contains approximately 30 nucleotides of SV40 DNA to the 5' end of the Hin G "L" strand transcript), there was clearly a strong preference for initiation of the RNA polymerase at the same site as was used in whole SV40 DNA. Therefore the entire promotor may well be contained in the ND3A segment of SV40, although the

possibility remains that the adeno virus DNA attached to SV40 DNA in Ad2 + ND3 contributed information to specify the promotor site. After the present sequence determination was completed we learned of the work of Allet and Roberts (25) who determined that E. coli RNA polymerase selectly binds to and protects the Hin GB junction in SV40 DNA. Thus if this site is part of the promotor it lies 30 nucleotides away from and preceeding the site where E. coli RNA polymerase actually initiates transcription of DNA.

The question arises as to why E. coli polymerase starts transcription exactly where it does on the DNA. One perhaps coincidental but striking feature of the sequences is that the polymerase begins transcription within a sequence of 12 nucleotides that is exactly repeated approximately 30 nucleotides further from the Hin G Hin B junction (8). Other than this there is no characteristic of this sequence that immediately suggests itself as neighboring sequence information specifying a polymerase start. An alternative possibility is that the polymerase has a "measuring function"and, for example, begins transcription at the first pyrimidine located at approximately 30 nucleotides from the mid point of a symmetric oligonucleotide sequence within the polymerase binding site (25, 26). Further analysis of polymerase binding sites and promotor sequences in other native DNA molecules should make it possible to deduce the common features recognized by polymerase and hence to determine how the initiation and transcription is encoded in DNA.

When cytoplasmic mRNA from SV40 infected cells was annealed to Ad2+ND3 DNA, eluted, and mapped, oligonucleotides T_1GL 9, 44 and 12 were prominent while oligonucleotide T_1BL-7 was not detectable(8). Therefore either termination of transcription, cleavage of precursor RNA or polyadenylation of the cleaved RNA appears to occur preferentially within oligonucleotide T_1BL-7, very close to or within the region

corresponding to the polymerase binding site on SV40 DNA. Portions of this short sequence therefore may be recognized by restriction endonucleases, bacterial RNA polymerase, and animal cell proteins concerned with termination of transcription or processing of RNA, a rather striking example of evolutionary conservation or convergence.

ACKNOWLEDGEMENT

We thank Frances Denoto, Su Fang Su, Jennifer Mathews, and Benjamin Jones for excellent technical assistance. This work was supported by Grant (ACS) VC1E and (NIH) CA10137-08.

REFERENCES

1 Losick, R. (1972) Ann. Rev. Biochem. 41, 409-446
2 Lebowitz, P. Weissman, S.M. & Radding, C.M. (1971) J. Biol. Chem. 246(16), 5120-5139
3 Dahlberg, J.E., J. Biol. Chem. (in press)
4 Paddock, G. & Abelson, J., (1973) Nature New Biol. 246(149), 2-6
5 Westphal, H. (1970) J. Mol. Biol. 50, 407-420
6 Zain, B.S., Dhar, R., Weissman, S.M., Lebowitz, P. and Lewis, A.M. Jr. (1973) J. Virol. 11 (5) 682-693
7 Danna, K. & Nathans, D. (1972) Proc. Nat. Acad. Sci. (U.S.A.) 69, 3097-3100
8 Dhar, R., Zain, B.S., Weissman, S.M., Pan, J., & Subramanian, K. (in press)
9 Dhar, R., Zain, B.S., Weissman, S.M., Lewis, A.M. Jr., PanJ. submitted for publication
10 Marotta, C.A., Levy, C.C., Weissman, S.M. & Varrichio, F. (1973) Biochemistry , 12, 2901-2904
11 Bernardi, G. in G. L. Cantoni and D . R. Davies (Editors) (1966) Procedures in Nucleic Acid Research, p. 37 Harper & Rowe, N.Y.
12 Smith, H.O., and Wilcox, K.W., (1970) J. Mol. Biol., 51, 379-391
13 Yoshimori, R.N. & Boyer, H.W. (1973) Ph.D. Thesis, University of California Medical Center, San Francisco
14 Marotta, C.A., Lebowitz, P., Dhar, R., Zain, B.S. and Weissman, S.M., Methods in Enzymology (Academic Press) Eds. Colowick, J.P., & Kaplan, N.D. (in press)
15 Subramanian, K.N.S., Zain, S. Weissman, S.M., Pan, J., Submitted for publication
16 Bigger, C.H., Murray, K., Murray, N.(1973) Nature New Biol. 244(131) 7-10
17 Boyer, H.W., Chow, L.T., Dugaiczyk, A., Hedgpeth, J. & Goodman, H.M. (1973) Nature New Biol. 244(132) 40-43

18 Lewis, A.M.Jr., Levine, A.S., Crumpacker, C.S., Levine, M.J., Samaha, R.J., & Henry, P.H. (1973) J. Virol. 11, 655-664
19 Kelly, T., Smith, H.O. (1970) J. Mol. Biol. 51, 393-410
20 Huang, E.S., Newbold, J.E., and Pagano, J.S. (1973) J. Virol. 11, 508-514
21 Lebowitz, P. personal communication
22 Jacob, F., Ullman, A. & Monod, J. (1964) CR Acad. Sci. 258, 3125-3128
23 Epstein, W., & Beckwith, J.R., (1968) Ann. Rev. Biochem. 37, 411-436
24 Blattner, F.R., & Dahlberg, J.E. (1972) Nature New Biol. 237, 227-232
25 Roberts, R., personal communication
26 Maniatis, T. & Ptashne, M. (1973) Nature, 246, 133-136
27 Brownlee, G. & Sanger, F. (1969) Europ. J. Biochem., 11, 395-408