Volume 2 number 11 November 1975 Nucleic Acids Research

Studies on bacteriophage fd DNA. III. Nucleotide sequence preceding the RNA start-site on a promoter-containing fragment.

K. Sugimoto, H. Sugisaki, T. Okamoto and M. Takanami

Institute for Chemical Research, Kyoto University, Uji, Kyoto, Japan

Received 8 September 1975

ABSTRACT

A short DNA fragment containing a strong promoter was isolated from phage fd replicative form DNA with the use of restriction endonucleases, and the sequence of 110 nucleotides in the region preceding the RNA start-site was determined. The sequence was : (5') CGGTCTGGTTCGCTTTGAGGCTCGAATTAAAACGCGATA-TTTGAAGTCTTTCGGGCTTCCTCTTAATCTTTTTGATGCAATTCGCTTTGCTTCTGACTATAATAGACAGG (3').

INTRODUCTION

In order to obtain information about the structure which specifies the initiation of transcription, attempts have been made to isolate DNA fragments containing promoters from phage fd replicative form(RF) DNA[1,2]. In previous work, we localized the promoter regions on the cleavage map of fd DNA, and found that one of the promoters for G-start RNA(named G2 promoter) is located on a fragment of about 230 base pairs long[2]. This fragment(named Hap-Hga V) is generated by digestion with two Haemophilus restriction endonucleases, Hap and Hga. RNA of about 120 bases long(named G2 RNA) is efficiently synthesised on this fragment, indicating that all the information needed for the promoter function is contained in this fragment. RNA polymerase binds tightly to this fragment at the RNA start-site, and protects about 45 base pairs from DNase-digestion. The sequences of the polymerase binding-site[3,4] and of G2 RNA[3] have been determined. Recent studies on the Lac and λ promoters however suggest that the promoter recognition site by RNA polymerase is separated from the site at which the polymerase forms the initiation complex[5,6,7]. In this communication, we determined the sequence of the entire region preceding the G2 RNA start-site of Hap-Hga V (non-transcribed region) in order to compare the sequence with those of other promoter regions. The non-transcribed region of Hap-Hga V was sequenced by the following procedures.

(A) On synthesising RNA at low salt, transcription in the reverse direction was initiated at an additional site, though RNA initiation at the G2 promoter was

still predominant. The site was located near the right end of the template, so that the resulting transcript covered nearly the entire region of the template (Fig.1 A). This (-)strand transcript(abbreviated (-)RNA·N) was used for the sequence analysis.

(B) Hap-Hga V was denatured by heating and incubated in an RNA synthesising mixture containing a primer oligonucleotide. Under the conditions used, transcription was initiated at many sites along each strand of the template with or without the primer. The synthesised products were fractionated by hybridisation to fd (+)strand DNA(phage DNA), and the (-)strand transcripts were collected. The (-)strand transcripts(abbreviated (-)RNA·HD) were further fractionated by gel electrophoresis and used for the sequence analysis(Fig.1 B).

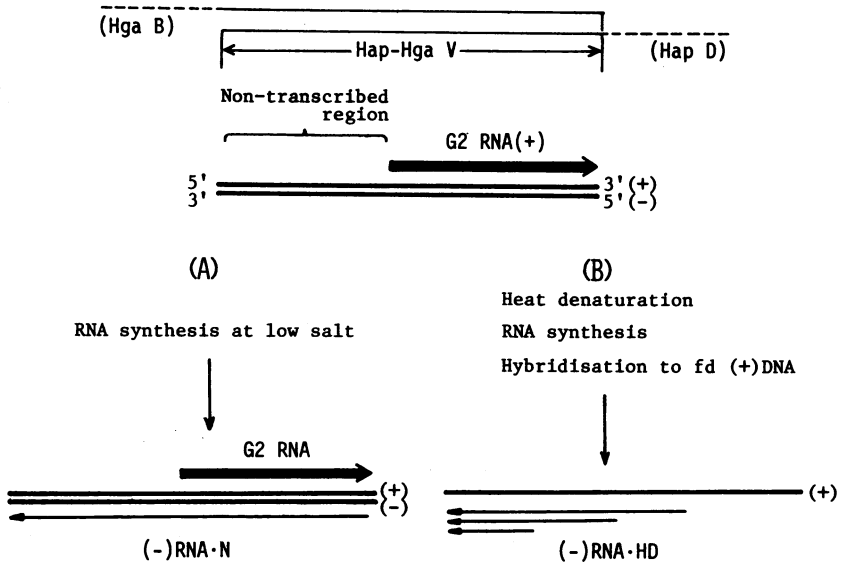

Fig.1 The procedures for sequencing the "non-transcribed" region of Hap-Hga V.

MATERIALS AND METHODS

Preparation of Hap-Hga V: As Hap-Hga V is yielded by digestion of the fragment Hap D with Hga[2], Hap D was first isolated from the Hap-digest of fd RF DNA by gel electrophoresis and then digested with Hga. Hap-Hga V was isolated from the digest by gel electrophoresis. The procedures for preparation of fd RF DNA and restriction endonucleases, Hap and Hga, and the conditions used for digestion of DNA and gel electrophoresis have been described previously[1].

Preparation of (-)RNA·N: The conditions used for RNA synthesis were essentially identical to those described previously[2], except that KCl was omitted from the

reaction mixture. The reaction mixture(1 ml) contained 8 mM $MgCl_2$, 40 mM tris (pH 7.9), 0.1 mM dithiothreitol, 50 μM of a nucleoside (α^{32}P)triphosphate(abbreviated NTP) and 0.2 mM each of three other NTPs, about 2 pmol of Hap-Hga V, and about 15 pmol of RNA polymerase. After incubation for 1 hr at 37°C, the synthesised products were isolated and electrophoresed for 3 hrs at 150V on 5% gel columns containing 7 M urea[3]. The resulting radioautograph is shown in Plate I(a) in comparison with RNA formed in the presence of 0.15 M KCl(Plate I b). On synthesising RNA at low salt, an additional band is seen at the indicated position. This RNA species only hybridised to fd (+)strand DNA, whereas G2 RNA hybridised to fd (-)strand DNA[2]. The size of RNA was estimated to be about 220 bases long from its mobility.

Preparation of (-)RNA·HD: Hap-Hga V(about 2 pmol) was heated for 5 min at 100°C, rapidly cooled, and added to an RNA synthesising mixture(1 ml) containing 8 mM $MgCl_2$, 50 mM KCl, 40 mM tris(pH 7.9), 0.1 mM dithiothreitol, 20 μM each of a (α^{32}P)NTP and three other NTPs, 4 μmol of AAAU, and about 15 pmol of RNA poly-

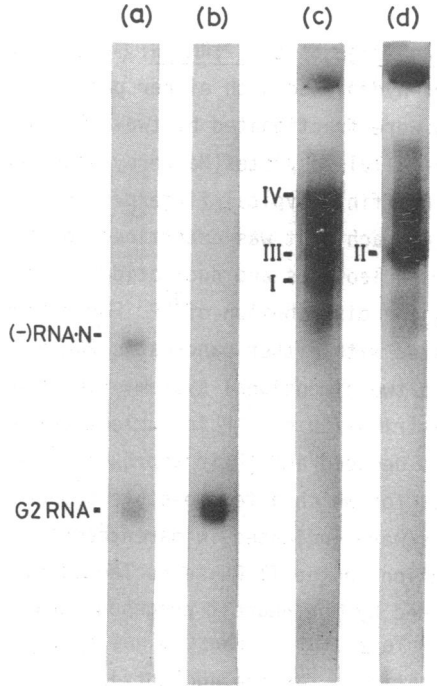

Plate I Polyacrylamide gel electrophoresis of RNA formed on Hap-Hga V.

(a) RNA formed on intact Hap-Hga V at low salt(5% urea-gel, 150V 3hrs). (b) RNA formed on intact Hap-Hga V at high salt(5% urea-gel, 150V 3 hrs). (c) (-)strand RNA formed on heat-denatured Hap-Hga V in the presence of AAAU(15% urea-gel,150V 3 hrs). (d) (-)strand RNA formed on exonuclease III-treated, heat-denatured Hap-Hga V in the presence of AAAU(15% urea-gel, 150V 3 hrs).

merase. In some experiments, Hap-Hga V was treated for 30 min at 37°C with 40 units of E.coli exonuclease III, before heat-denaturation. Incubation was done for 6 hrs at 37°C, and terminated by shaking with 80% phenol. After removal of phenol with ethylether, 25 µg of freshly dissolved DNase I were added, and the solution was incubated for 5 min at 37°C to digest the template. The solution was then treated with 80% phenol, and passed through a Sephadex G50 column. The RNA fraction obtained was mixed with 1 A_{260} unit of fd phage DNA in 0.9 M NaCl-0.09 M Na-citrate. The mixture was heated for 5 min at 100°C, and then held for 3 hrs at 51°C. After cooling, the mixture was passed through a nitrocellulose filter(Millipore HA 0.45 µ). The filter was washed with 0.3 M NaCl-0.03 M Na-citrate. (^{32}P)RNA retained on the filter was dissociated by heating for 5 min at 100°C in 20 mM tris(pH 7.6). The dissociated RNA was concentrated and electrophoresed for 3 hrs at 150V on 15% gel columns containing 7 M urea[3]. Typical radioautographs obtained are shown in Plate I (c),(d). RNA was extracted from the band regions, re-electrophoresed on 15% gel, and used for the sequence analysis.

RESULTS AND DISCUSSION

Analysis of pancreatic and T1 RNase-digests of (-)RNA·N: (-)RNA·N was synthesised with each of four (α^{32}P)NTPs, and hydrolysed with either pancreatic RNase or T1 RNase. The resulting nucleotides were fractionated by two-dimensional chromatography on polyethyleneimine(PEI)-cellulose plates(Macherey-Nagel Co.) according to the method of Mirzabekov and Griffin[8]. Typical fingerprints are shown in Plate II (a),(b). Distribution of ^{32}P in each spot was determined to estimate the molar yields. The sequences of shorter nucleotides and nucleotides adjacent to their 3'-end were simply determined by the distribution of ^{32}P. For analysis of longer nucleotides, each spot was digested with either pancreatic RNase, T1 RNase or U2 RNase, and chromatographed by the two-dimensional system. Each spot yielded was then digested by T2 RNase and distribution of ^{32}P in nucleotides was determined[3]. The sequences of nucleotides thus deduced and their approximate molar ratios are shown in Table I. The nucleotides for which a few possible sequences were deduced by the nearest neighbour analysis were indicated in parentheses. The sequence of P24 was deduced from the information on the T1 RNase-nucleotides, though two alternative sequences were obtained by the nearest neighbour analysis. The yield of T26 was always less than one mole in the T1 RNase-digest(see Table I). The starting nucleotides were not identified, as the nucleotides containing terminal triphosphates stayed at the origin under the conditions used for chromatography.

Analysis of T1 RNase-digests of (-)RNA·HD: (-)RNA·HD labeled with a single (α^{32}P)

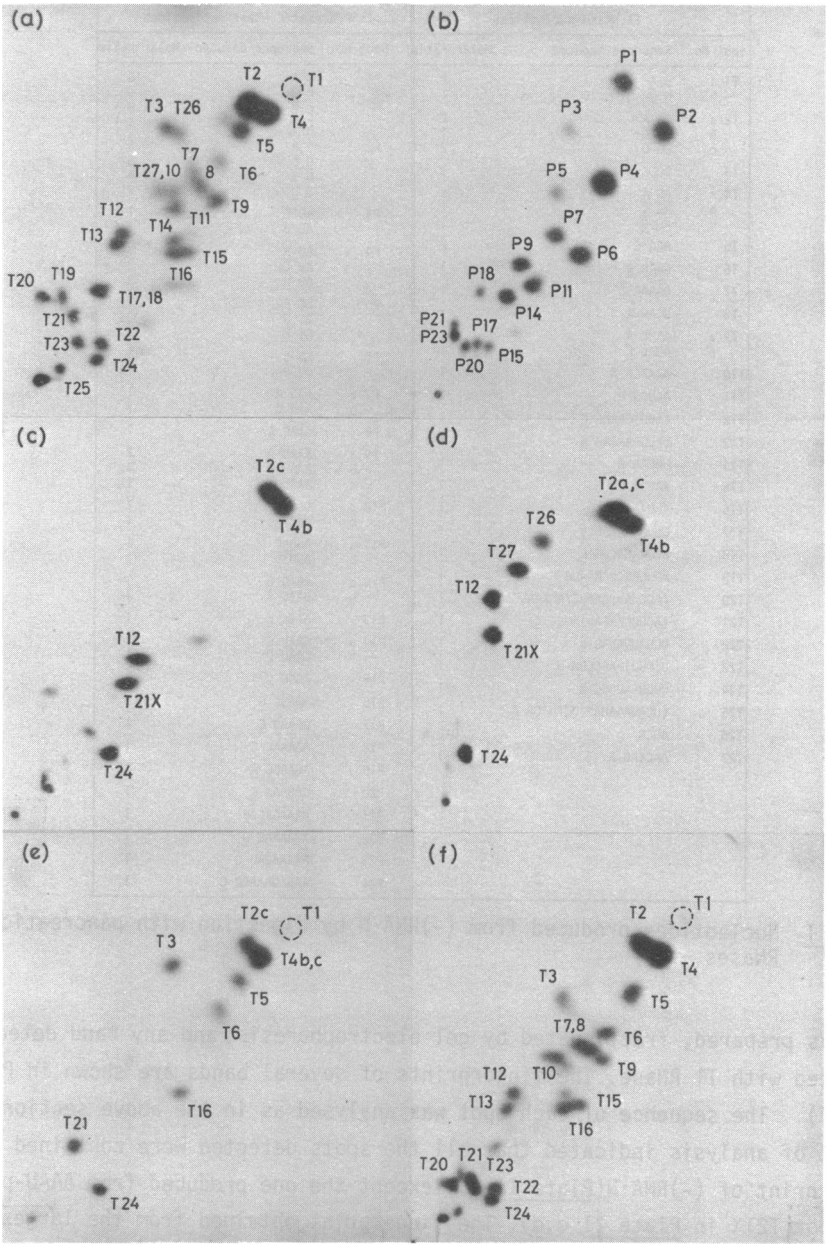

Plate II **Two-dimensional fingerprints of pancreatic and T1 RNase-digests of RNA synthesised on Hap-Hga V.**

(a) T1 RNase-digest of (-)RNA·N. (b) Pancreatic RNase-digest of (-)RNA·N. (c)-(f) T1 RNase-digests of band I to IV in Plate I c,d. The T1 RNase-digests have been labelled with (α^{32}P)GTP and the pancreatic RNase-digest with (α^{32}P) UTP. The sequence of each spot is given in Table I.

T1 RNase-nucleotides				Pancreatic RNase-nucleotides		
Spot No.	Sequence deduced	Molar ratio	Spot No.	Sequence deduced	Molar ratio
T1 a	G A	2	P1	C A	>10
b	G U	1		C C	≈ 6
T2 a	CG A	2		C G	≈ 5
b	CG G	1		C U	≈ 5
c	CG U	2	P2	U A	≈ 6
T3	CCCG A	1		U C	≈ 6
T4 a	AG A	1		U G	≈ 4
b	AG C	1		U U	≈ 7
c	AG G	1	P3	AC C	1
T5	AAG C	2		AC U	1
T6	AAAG A	1	P4	AU A	3
T7	CAAAG C	1		AU C	2
T8	UCAG A	1		AU U	3
T9 a	AAUG A	1	P5	GC A	1
b	AUAG C	1		GC G	2
T10	AAAACG A	1		GC U	1
T11	AAUCG U	1	P6	GU C	3
T12	C(CU)CAAAG C	1		GU U	1
T13	CAUCAAAAAG A	1	P7 a	AAU A	1
T14	UUCAG A	1	b	AAU U	1
T15	AAUUG C	1	P8	AAAC A	1
T16	AUUAAG A	1	P9 a	AAAU A	2
T17	CUUUAAACAG U	1	b	AAAU C	1
T18	UCCAAUACUG C	1	c	AAAU G	1
T19	AAUCCCCUCAAAUG C	1	P10 a	GAC C	1
T20	(ACC/AU/AAAUC/AAAAAUC)AG G	1	b	AGC G	1
T21	(ACUUC/AAAU)AUCG C	1	P11 a	AGU C	1
T22	UCUAUUAUAG U	1	b	AGU U	1
T23	U(CUU)UACCCUG U	1	P12 a	AGAC C	1
T24	UUUUAAUUCG A	1	b	GAAC C	1
T25	(UCAUAAAUAU/UCAU)UG A	1	P13	GAGC C	1
T26	ACCG	0-0.6	P14 a	GAAU C	1
T27	AACCAG A	1	b	GAAU U	1
			P15	AGGU C	1
			P16	AAAGC G	2
			P17	GGAAU C	1
			P18	AAAAAU C	1
			P19	AGAAGC A	1
			P20	GAGAAU G	1
			P21	GAAAGAC U	1
			P22	AGAAAAC G	1
			P23	AAAAAGAU U	1
			P24	AAGAGGAAGC C	1

Table I Nucleotides produced from (-)RNA·N by digestion with pancreatic and T1 RNases.

NTP was prepared, fractionated by gel electrophoresis, and any band detected was digested with T1 RNase. The fingerprints of several bands are shown in Plate II (c)-(f). The sequence of each spot was analysed as in the above section. The result of analysis indicated that all the spots detected were contained in the fingerprint of (-)RNA·N(Plate II a), except the one produced from AAAU-primed RNA(spot T21X in Plate II c,d). The fingerprint obtained from the largest (-)RNA·HD was very similar to that of (-)RNA·N(Plate II f). The spot number of the major nucleotides detected are indicated on the fingerprint.

Ordering the T1 RNase-nucleotides: The sequences of G2 RNA and of the RNA polymerase-binding site on Hap-Hga V have already been determined[3]. The total length of the determined sequence represents about 60% the length of Hap-Hga V. In Fig.2 (A

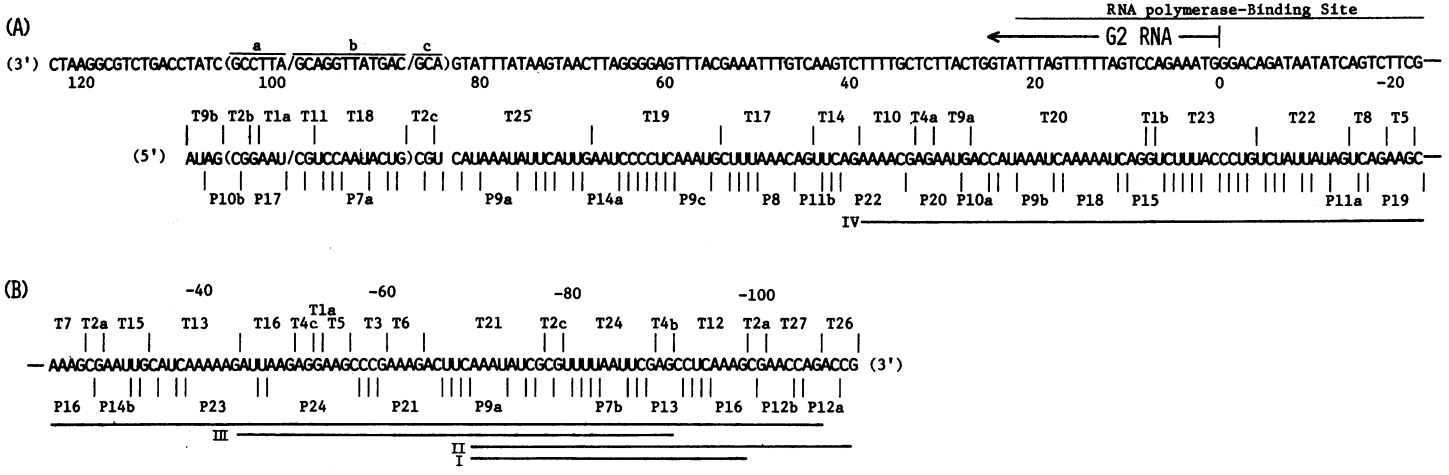

Fig.2 Arrangment of pancreatic and T1 RNase-nucleotides produced from (-)RNA·N. The regions covered by (-)RNA·HD are underlined. Roman numerals represent the bands indicated by the side of Plate I c,d.

the sequence was indicated as the sequence of DNA in the opposite direction (3'← 5') for convenience. On the other hand, the size of (-)RNA·N has been estimated to be about 220 bases long, which corresponds to about 95% the length of the template. Therefore, the sequence of (-)RNA·N should cover the polymerase-binding site and the most part of G2 RNA. On the basis of this assumption, we aimed to arrange the T1 RNase-nucleotides of (-)RNA·N on the sequence previously determined, and nineteen nucleotides were successfully ordered as indicated in Fig.2 (A). The products expected from the 109th to 123th position were not detected. Thus, we concluded that (-)RNA·N was initiated at the 109th position. In the previous analysis[3], the order of three segments between the 84th and 104th positions had not been determined(Fig.2 A, segments a, b, and c). In the present analysis, T25 just fits in the region between the 69th to 84th position. Thus, the segment next to the 83th position should be c. Although the order of two segments, a and b, is still not known, the same pancreatic and T1 RNase-products should be yielded from either arrangment. Eighteen T1 RNase-nucleotides remaining were assigned to the right of the RNA polymerase-binding site. These nucleotides were ordered by the analysis of shorter (-)RNA·HD.

The T1 RNase-nucleotides yielded from three different sizes of (-)RNA·HD(bands I to III in Plate I c,d) are shown in Plate II(c)-(e). Bands I and II yielded a new spot T21X. This spot was found to contain the primer at the 5'-end, and the sequence was deduced to be AAAUAUCG. Among the T1 RNase-nucleotides produced from (-)RNA·N, one of the sequences deduced for T21(ACUUCAAAUAUCG or AAAUACUUCAUCG) contained the sequence of T21X at the 3'-end. Thus, we concluded that bands I and II had been initiated with the primer within the T21 region. In conjunction with the information on the pancreatic RNase-products, the sequence of T21 was also deduced to be ACUUCAAAUAUCG. Band I produced only five T1 RNase-nucleotides. Band II contained all the T1 RNase-nucleotides of band I and three additional nucleotides, T2a, T26, and T27. Bands I and III overlapped with T2c, T4b, T21, and T24. With these observations and the information on nucleotides linked to G, the T1 RNase-nucleotides in the region covered by bands I and II were ordered as in Fig.2 (B). It was also concluded that the sequence of (-)RNA·N is terminated by T26, as T1 RNase-digestion of (-)RNA·N always yielded less than one mole of T26(see Table I). This order has been confirmed by analysis of the pancreatic RNase-products of bands I and II(data not shown). Two alternative sequences have been obtained for T12(see Table I). As pancreatic RNase-digestion of band I yielded P13, the sequence of T12 was deduced to be CCUCAAAG.

The order of the T1 RNase-nucleotides produced from the left part of band III was determined by analysis of the pancreatic RNase-nucleotides. Among the pan-

creatic RNase-nucleotides of (-)RNA·N, the sequence of P24 was AAGAGGAAGC(C). T1 RNase-digestion of this nucleotide yields T1a, T4c, and T5 from the inside of the chain and AAG(A) from the 5'-end. (Py)AAG(A) is only contained in T16. On the other hand, P21 contains the sequence of T6. On the basis of these observations, the order of the T1 RNase-nucleotides in the left part of band III was determined to be T16:T4c:T1a:T5:T3:T6.

T1 RNase-nucleotides remaining are T2a, T7, T13, and T15. These nucleotides should be located between T5 and T16 in the sequence of Fig.2. The order was determined to be T7:T2a:T15:T13, because three long pancreatic RNase-nucleotides remaining(P14b, P16, and P23) are only produced by this arrangment. As indicated in Fig.2, all the oligonucleotides produced by pancreatic RNase-digestion of (-)RNA·N can be perfectly arranged in the sequence of (-)RNA·N finally deduced. In addition, this sequence has further been confirmed by analysis of other (-)RNA·HD and partial T1 RNase-digests of (-)RNA·N(data not shown).

In Fig.3 (A), the total sequence in the region preceding the G2 RNA start-site is shown as the sequence of DNA. All the information essential to the promoter function should be contained in this sequence. As indicated in the figure, the sequence contains a few regions with two-fold rotational symmetry. The CTTT sequence appears to occur frequently. As has been noted in the Lac promoter[7], the fd promoter region appears to contain a high AT-rich region flanked with the regions of relatively high GC-content(Fig.3 A). By comparison with the sequences of other promoter regions[5,6,7,9,10,11], the fd sequence bears some resemblance

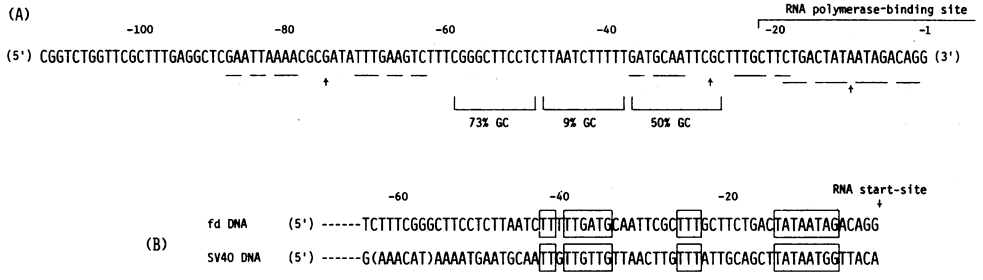

Fig.3

(A) Total sequence of the region preceding the G2 RNA start-site on Hap-Hga V.
(B) Sequence homology in the promoter regions on fd and SV40 DNA(The sequence of SV40 DNA was taken from ref.10).
In (A) regions with two-fold rotational symmetry are underlined (↑: axis of symmetry). In (B) homologous sequences are boxed.

to the sequence preceding an E.coli RNA polymerase initiation site on SV40 DNA[10]. The homologous sequences in the corresponding region are indicated in Fig.3 (B) for comparison. As has been noted by Pribnow[11], both the sequences contain TATAAT in the RNA polymerase binding-site. However, it appears to be still premature to discuss about the sequence features of the promoter. It would be necessary to accumulate more information about other promoter sequences.

It should be mentioned that RNA synthesis was initiated at an additional site at low salt. The site was located at about 15 base pairs inside the leftmost end of Hap-Hga V (Hga cleavage site, see Fig.1). No RNA chain was initiated at the corresponding site on Hap D at low salt. We assume that the cleavage of Hap D with Hga induces a specific configuration at the terminal region which permits the binding of RNA polymerase at low salt.

REFERENCES

1. Takanami,M., Okamoto,T., Sugimoto,K., and Sugisaki,H.(1975) J. Mol. Biol., 95, 21-32.
2. Okamoto,T., Sugimoto,K., Sugisaki,H., and Takanami,M.(1975) J. Mol. Biol., 95, 33-44.
3. Sugimoto,K., Okamoto,T., Sugisaki,H., and Takanami,M.(1975) Nature, 253, 410-414.
4. Schaller,H., Gray,C., and Herrmann,K.(1975) Proc. Natl. Acad. Sci. U.S., 72, 737-741.
5. Maniatis,T., Ptashne,M., Barrell,B.G., and Donelson,J.(1974) Nature, 250,394-397.
6. Walz,A. and Pirrotta,V.(1975) Nature, 254, 118-121.
7. Dickson,R.C., Abelson,J., Barnes,W.M., and Reznikoff,W.S.(1975) Science, 187, 27-35.
8. Mirzabekov,A.D. and Griffin,B.E.(1972) J. Mol. Biol., 72, 633-643.
9. Sekiya,T. and Khorana,H.G.(1974) Proc. Natl. Acad. Sci. U.S., 71, 2978-2982.
10. Dhar,R., Weissman,S.M., Zain,B.S., Pan,J., and Lewis,A.M.(1974) Nucleic Acid Research, 1, 595-613.
11. Pribnow,D.(1975) Proc. Natl. Acad. Sci. U.S., 72, 784-788.