Volume 4 Number 4 April 1977 Nucleic Acids Research

Nucleotide sequence of the SV40 DNA restriction fragment Hind C - Hap 2

H. Van Heuverswyn, A. Van de Voorde and W. Fiers

Laboratory of Molecular Biology, University of Ghent, Ledeganckstraat, 35
9000 Gent, Belgium

Received 2 March 1977

ABSTRACT

We here report the nucleotide sequence of the SV40 DNA fragment Hind C - Hap 2. The fragment was labeled at the 5'-ends by means of polynucleotide kinase and γ-(^{32}P)ATP and digested with a suitable restriction enzyme. The separated products were then partially degraded with the base-specific reagents dimethylsulphate or hydrazine followed by direct analysis on polyacrylamide gel. The Hind C - Hap 2 sequence is 126 base pairs long and one of the three possible reading frames for translation does not contain any termination codon. So, although no protein is known to be encoded by this region, the possibility cannot yet be completely ruled out. The sequence also contains several AT-rich blocks.

INTRODUCTION

SV40 DNA is a double-stranded, supercoiled, circular molecule containing approximately 5200 base pairs. It is cleaved by Hind (II + III) restriction enzyme into 13 fragments, 11 of which were ordered by Danna and Nathans (1). The two minor fragments Hind L and Hind M were subsequently discovered and mapped between 0.747 and 0.757 as part of the late region by Yang et al. (2). We recently also reported their nucleotide sequence (3). Part of the sequence of the Hind C fragment has been determined by Subramanian et al. (4). Here we report the nucleotide sequence of the Hind C - P 2 fragment (footnote), generated by digestion of the Hind C fragment with the Hap (isoschizomer of Hpa II) restriction enzyme (H.aphirophilus). The biological information

Footnote : Hind C-fragment is cleaved by Hap-enzyme in two fragments, Hind C - Hap 1 and Hind C - Hap 2 (1)(2); the latter is abbreviated to Hind C - P 2 and is the subject of this paper. Fragments obtained by secondary digestion with Alu, Hae III or Hind II + III enzyme are indicated by A, E or H respectively.

encoded by this 126 base pairs long region is not known. However, Mertz et al. (5) and Shenk et al. (6) could obtain viable mutants with deletions encompassing the Hap-site (position 0.724). These results suggest that this region is non-essential, although the mutants with deletions extending towards position 0.74 on the standard map (i.e. near the right end of fragment Hind C - P 2 which maps at 0.747) grew poorly. Also insertion of a poly(dA.dT)-segment at the Hap (Hpa II)-site does not inactivate the virus, but again it does decrease the biological fitness (7).

MATERIALS

γ-(^{32}P)-ATP was prepared according to the method of Glynn and Chappell (8). Bacterial alkaline phosphatase and snake venom phosphodiesterase were purchased from Worthington Biochemical Corp. T_4 polynucleotide kinase was obtained from P.L. Biochemicals and pancreatic DNase from Boehringer Co. Dimethylsulphate was from Aldrich Co and hydrazine from Eastman Kodak Co.

METHODS

SV40 DNA and restriction fragments

All procedures for the preparation of SV40 DNA (strain 776) and the digestion conditions for the preparation of restriction enzyme fragments have been described elsewhere (9).

5' Terminal labeling of restriction fragments

Labeling of the 5' ends of the DNA fragments was done mainly according to Maxam and Gilbert (10) (cf. also Van de Voorde et al. (11), except that in some cases digestion with the restriction enzyme was carried out after dephosphorylation but before the polynucleotide kinase step (F. Thys and R. Contreras, personal communication). In general, the enzymatic reaction mixture was applied directly on a two-step gel (6 % - 12 % polyacrylamide) for the separation of the fragments, each labeled at only one 5' end.

Chemical degradation of the 5'-labeled fragments

Partial chemical degradation was performed according to the method of Maxam and Gilbert (10), of which only the principle is given here. The fragment is incubated with a suitable amount of dimethylsulphate (guanine and adenine specific cleavage) or hy-

drazine with or without salt (cytosine and thymine-cytosine specific cleavage respectively). In recent experiments the adenine specific degradation was also obtained by heating the DNA in 1.2 M NaOH. After chemical modification, strand scission is obtained by heating the treated DNA in piperidine. The resulting reaction mixture is then analysed on a 0.2 x 30 x 90 cm 20 % polyacrylamide slab gel run in a buffer which contains 7 M urea, 50 mM Tris-citrate (pH 8.0) and 5 mM EDTA.

Analysis of the 5'-terminal nucleotide of kinated fragments

In order to determine the 5'-terminal nucleotide, a small amount of the fragment labeled at one end was precipitated together with 35 µg of yeast RNA as carrier. The pellet was dissolved in a reaction buffer containing 10 mM Tris-chloride, pH 7.4, 5 mM $MgCl_2$ and bovine serum albumin (1 mg/ml). Subsequently, 0.1 ng pancreatic DNase and 0.5 µg snake venom exonuclease were added. After overnight incubation, the mixture was applied directly on a polyethyleneimine plate (20 x 20 cm, Polygram Cel 300 PEI-Macherey & Nagel & Co) which was developed by ascending chromatography in 0.2 M LiOH, pH 3.0, (12) as described before .

RESULTS AND DISCUSSION

The Hind C-fragment can be cleaved by the Hap (H.aphirophilus) or the Hpa II (H.parainfluenzae) enzyme in a 7.7 % and a 2.45 % fragment (1). This paper deals with the smaller fragment, referred to as Hind C - P 2. It can be further cleaved by a number of different restriction enzymes such as Eco RII, Hae III, and Alu. (The sequence also turned out to contain the putative recognition site of the Mbo II enzyme, but this splitting was not further checked). Previous results from our laboratory (13) have shown that the fragment is cleaved by the Alu-restriction enzyme into three smaller fragments Hind C - P 2 - A 1, Hind C - P 2 - A 2 and Hind C - P 2 - A 3, corresponding to 1.15 %, 0.76 % and 0.54 % of the SV40 DNA respectively. After 5'-labeling with T_4 polynucleotide kinase and γ-(^{32}P)-ATP and subsequent Alu digestion, only the two larger fragments are labeled.

In one experiment, the Hind C - P 2 fragment was split by the Alu enzyme in mainly 2 pieces, about 1.3 % and 1.15 % respectively, and undoubtedly generated by a partial cleavage at the

Alu-site closest to the Hpa II site (fig. 1). The partial chemical degradation pattern of the 1.3 % fragment (band 2 in fig. 1) is shown in figure 2a. The sequence of 63 consecutive nucleotides can easily be deduced from this single gel. By experience we know that in this system a pentanucleotide runs slightly faster than the fluoresceine dye marker. So, the fastest moving band seen on the gel presumably corresponds to the loss of the C-residue, which forms part of the Hind II site. This was further proven by analysis of the same fragment on a gel run for a shorter time (not shown). As the 5'-end cannot be identified by the Maxam-Gilbert degradation procedure, it was determined by direct analysis as described in Materials and Methods. The only labeled 5'-nucleotide was pA (fig. 3). It may be noted that following the AAAAG sequence there is a rather peculiar region where a few nucleotides are missing. In a later experiment this turned out to be precisely the Alu-site which had not been cleaved. This can be seen in figure 2b which represents a longer run of the same strand of the Hind C - P 2 fragment, but cleaved by the Hae III restriction enzyme. From this pattern the sequence up to nucleotide 49 (i.e. 77 nucleotides starting from the Hind II + III recognition site) can unambiguously be deduced. We have no good explanation why in the experiment shown in figure 2a some nucleotides constituting the Alu-site have escaped chemical reaction (all proteins had been removed by prior phenolization).

The remaining sequence of the Hind C - P 2 fragment was determined on the other strand, starting from the Hpa II-site (fragment Hind C - P 2 - A 1; band 3, fig. 1). Its partial chemical degradation patterns is shown in figure 4a and the sequence of the first 32 nucleotides can be deduced from the gel patterns. As expected, the 5'-terminal nucleotide was indeed pC (fig. 3). In order to extend the sequence further, the fragment Hae F (9) was terminally labeled and digested with Hind II + III. The second largest fragment, Hae F - H 2, is derived from the Hind C - P 2 region and its chemical degradation allowed to establish the sequence up to nucleotide 70 (figure 4b and 4c). The combined data lead unambiguously to the complete sequence of the Hind C - P 2 fragment (figure 5) and the terminal sequences are in agreement with the known recognition sites of the restriction enzymes

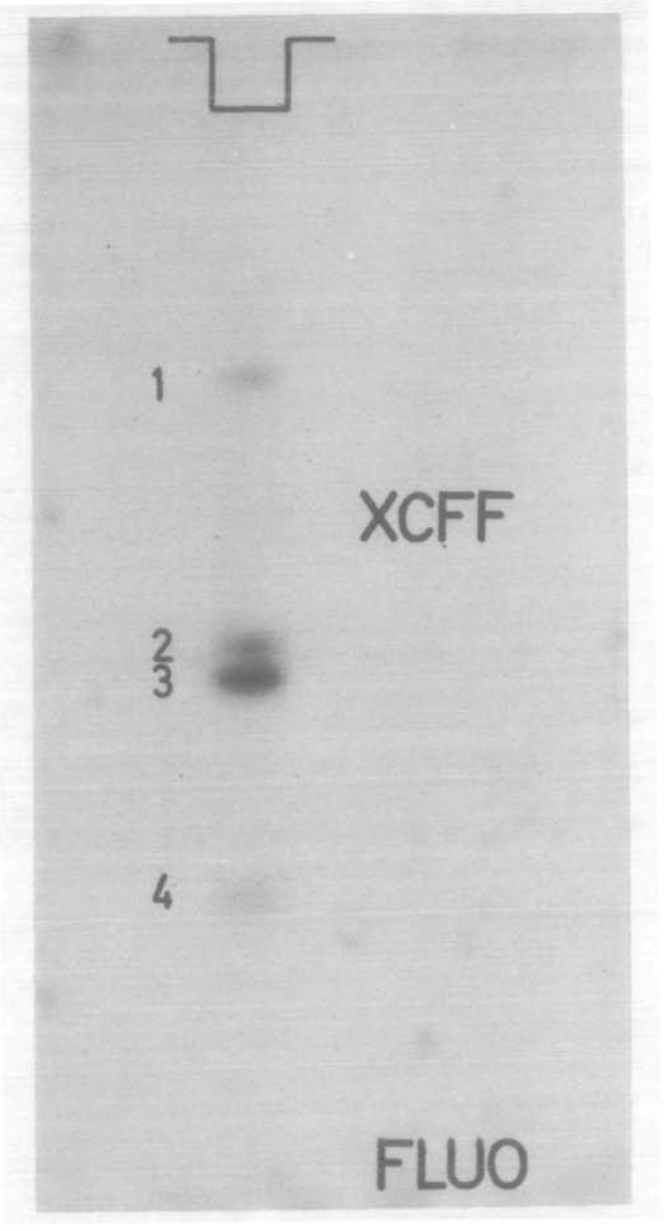

Figure 1 : Separation of the different fragments generated by digestion of the Hind C - P 2 fragment with Alu restriction endonuclease. After 5'-labeling with T_4 polynucleotide kinase and γ-(^{32}P)-ATP and subsequent Alu digestion, the reaction mixture was applied directly on an 8 % polyacrylamide slab gel and run at 200 V overnight. Bands 1, 2, 3 and 4 represent Hind C - P 2 (intact fragment), Hind C - P 2 - A 2 + 3, Hind C - P 2 - A 1 and Hind C - P 2 - A 2 respectively.

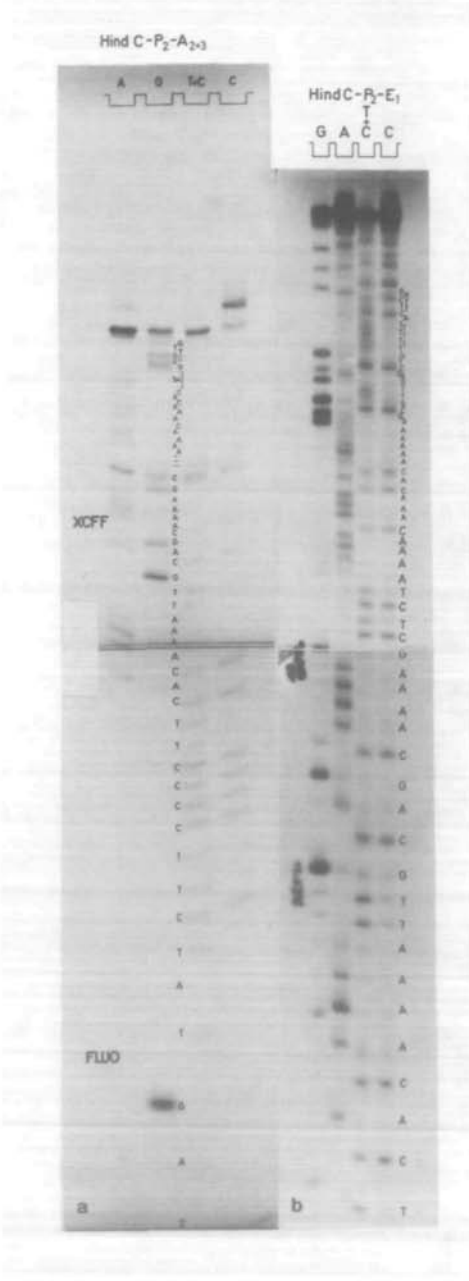

Figure 2 (a) : Sequence analysis of Hind C - P 2 - A 2 + 3 (band 2 in figure 1) by the chemical degradation procedure. The double-stranded, 5'-labeled DNA fragment was mixed with carrier DNA and incubated with dimethysulphate (G and A) or with hydrazine (C and T + C). The reaction was stopped by precipitation with ethanol. Elimination of the reacted bases and chain cleavage was obtained by heat-alkali treatment. The fragments were fractionated on a 20% acrylamide gel, containing 7 M urea at pH 8.0, as described under Methods. XCFF and FLUO indicate the positions of the xylene cyanol FF and fluoresceine dye-markers respectively.

Figure 2 (b) : Sequence analysis of Hind C - P 2 - E 1 fragment, 5'-terminally labeled at the Hind II + III restriction site, by the chemical degradation procedure. The upper part has been omitted from the figure (the distance from the origin to the undegraded bands was 20 cm).

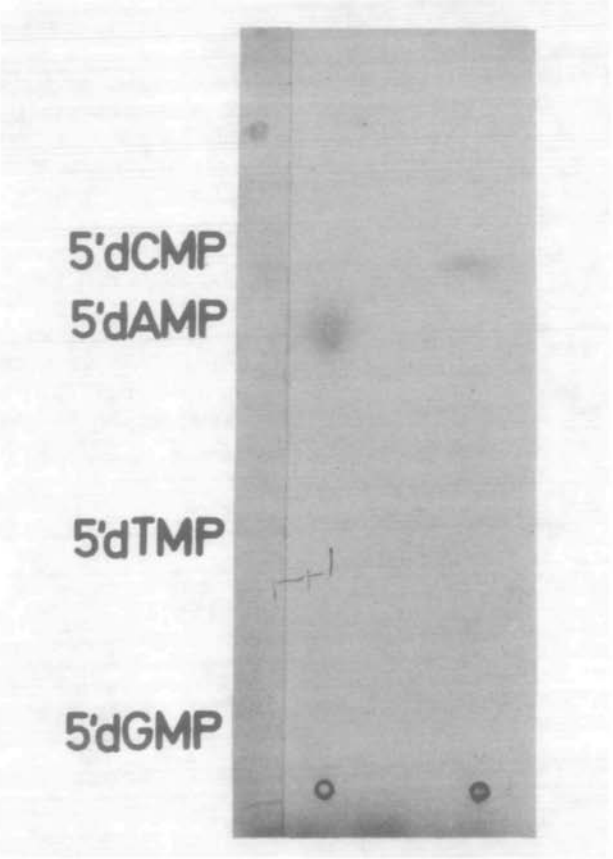

Figure 3 : Identification of the 5'-terminal nucleotide of each strand of the restriction fragment Hind C - P 2. The 5'-terminally labeled DNA strands were totally digested with snake venom phosphodiesterase and pancreatic DNase and the nucleotides were chromatographed on a polyethyleneimine thin layer plate at pH 3.0, as described under Methods. Right, the Hind C - P 2 - A 1 and left the Hind C - P 2 - A 2 + 3 5'-terminal nucleotide.

Hind II (14) and Hap (or Hpa II) (15).

As mentioned in the introduction, we recently reported the nucleotide sequence of the Hind L and Hind M fragments (3), which immediately follow the sequence presented here. Together, they represent a stretch of 180 nucleotides. Assuming for a moment that it is reasonable to interpret this region in terms of translation, it is noteworthy that there is one possible reading frame, which runs clockwise (in the direction of late transcription) through Hind C - P 2 (figure 5) and continues through Hind L and Hind M until it ends with a UAA-signal at the right end of Hind M (3).

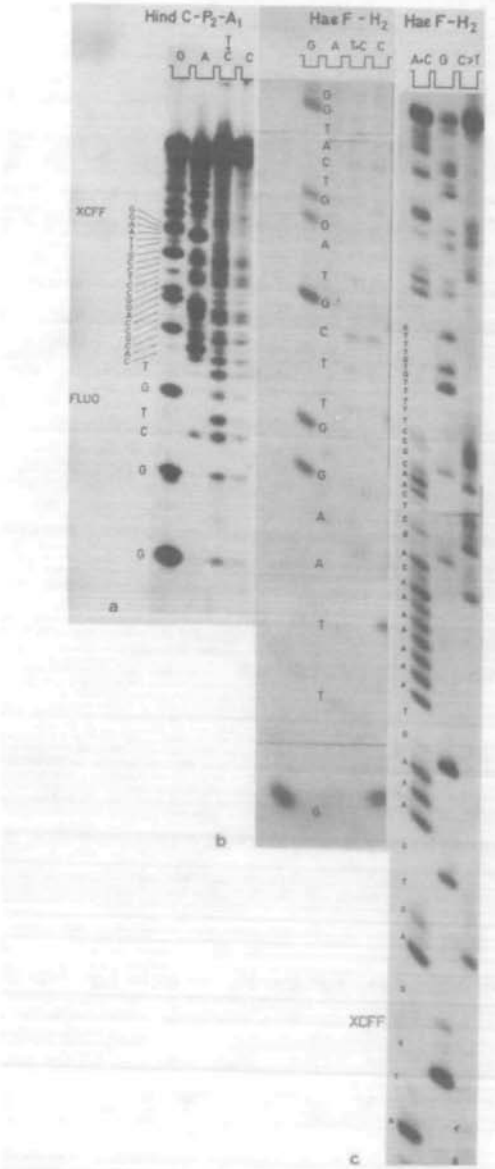

Figure 4 : (a) Sequence analysis of Hind C - P 2 - A (band 3, figure 1) and (b + c) of the Hae F - H 2 fragment by the chemical degradation procedure. This fragment was generated by 5'-terminal labeling of the Hae F fragment and subsequent digestion with the Hind II + III enzyme. (The notation Hae F-H 2 indicates how this fragment was generated; it is identical, however, to the fragment Hind C - P 2 - E 1 (9). In figure 4b, the upper part has been omitted from the figure (the distance from the origin to the upper G-band was 47 cm). (The spots on the very right of the figure are due to an adjacent lane belonging to the chemical degradation of another fragment). In fig. 4c, which represents a longer run from the Hae F - H 2 fragment, the distance from the origin to the undegraded bands was 25 cm.

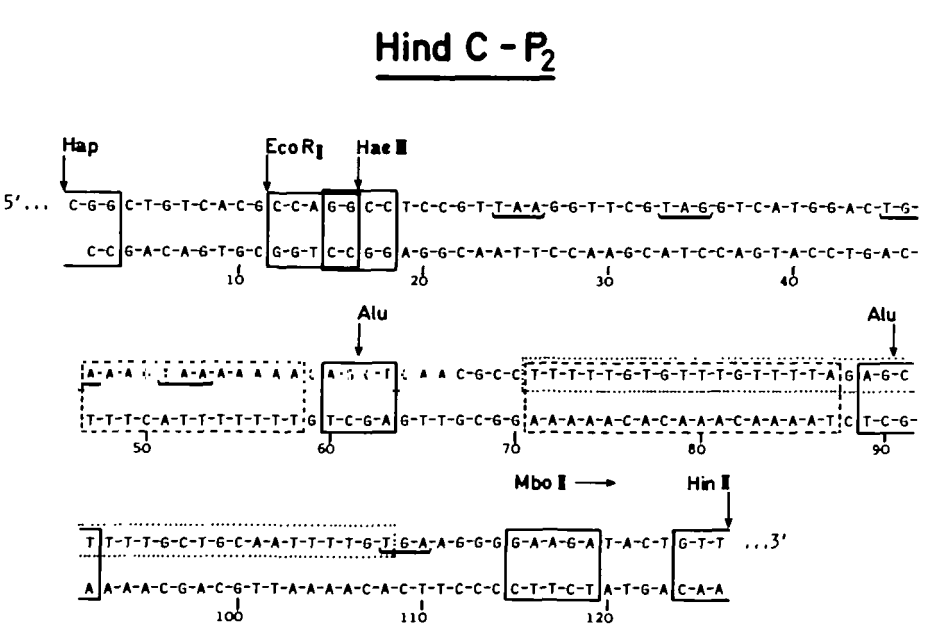

Figure 5 : The nucleotide sequence of the restriction fragment Hind C - P 2. The recognition sites of several restriction endonucleases are indicated in full boxes. The dashed boxes indicate AT-rich blocks and the dotted boxes indicate A-rich regions and T-rich regions on the same strand. The upper strand has the same polarity as late mRNA; two of the three possible reading frames for translation contain termination codons (underlined). Whether this region is in fact translated is not known.

More data, however, either biochemical or chemical, will be required before it will be known if this sequence does indeed code for a polypeptide. If so, it would presumably be a non-essential protein in view of the viable deletion mutants around the Hap (Hpa II)-site, isolated by Mertz et al. (5), by Shenk et al. (6) and by Carbon et al. (7).

Another remarkable feature of the sequence as here reported are the two AT-rich blocks (figure 5). The first is 12 basepairs in length and contains 92 % AT, while the second extends over 17 base pairs and contains 88 % AT. Furthermore in each block the A-residues are predominantly in one strand and the T-residues in the other. In fact, considering only the upper strand in Fig. 5, there is a 20 residues long region which contains 65 % A-residues, a spacer of 4 G- or C-residues, and a second region of 38 residues,

61 % of which are T. Similar asymmetric AT-rich blocks have previously been shown to be present in the neighbouring Hind L and Hind M fragment (3). Whether such blocks have a special role as part of regulatory elements or recognition sites for specific interaction with proteins is at present not known.

Recently we learned that S. Weissman and coworkers have obtained the same sequence as here reported (16).

ACKNOWLEDGEMENT

We are grateful to Dr. A. Maxam and W. Gilbert for kindly making the details of their sequencing method available to us before publication. J. Vanderheyden contributed expert technical help. We thank the "Kankerfonds" of the Algemene Spaar- en Lijfrentekas of Belgium for financial support.

REFERENCES

1. Danna, K.J., Sack, G.H. and Nathans, D. (1973) J. Mol. Biol. 78, 363-376.
2. Yang, R.C.-A, Danna, K.J., Van de Voorde, A. and Fiers, W. (1975) Virology 68, 260-265.
3. Ysebaert, M., Thys, F., Van de Voorde, A. and Fiers, W. (1976) Nucleic Acids Res. 3, 3409-3421.
4. Subramanian, K.N., Dhar, R. and Weissman, S.M. (1977) J. Biol. Chem. 252, 355-367.
5. Mertz, J.E. and Berg, P. (1974) Proc. Nat. Acad. Sci. U.S.A. 71, 4879-4883.
6. Shenk, T.E., Carbon, J. and Berg, P. (1976) J. Virol. 18, 664 - 671.
7. Carbon, J., Shenk, T.E. and Berg, P. (1975) J. Mol. Biol. 98, 1-15.
8. Glynn, I.M. and Chappell, J.B. (1964) Biochem.J. 90, 147-149.
9. Yang, R.C.-A, Van de Voorde, A. and Fiers, W. (1976) Eur. J. Biochem. 61, 101-117.
10. Maxam, A.M. and Gilbert, W., Proc. Nat. Acad. Sci. U.S.A. (in press).
11. Van de Voorde, A., Contreras, R., Rogiers, R. and Fiers, W. (1976) Cell 9, 117-120.
12. Volckaert, G. and Fiers, W. (in preparation).
13. Yang, R.C.-A, Van de Voorde, A. and Fiers, W. (1976) Eur. J. Biochem. 61, 119-138.
14. Kelly, T.J. and Smith, H.O. (1970) J. Mol. Biol. 51, 393-409.
15. Garfin, D.E. and Goodman, H.M. (1974) Biochem. Biophys. Res. Comm. 59, 108-117.
16. Dhar, R., Subramanian, K.N., Pan, J. and Weissman, S. (1977) Proc. Nat. Acad. Sci. U.S.A. (in press).