Volume 9 Number 16 1981 Nucleic Acids Research

Promotion of specific in vitro transcription by excised "TATA" box sequences inserted in a foreign nucleotide environment

P.Sassone-Corsi, J.Corden, C.Kédinger and P.Chambon

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS et Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Faculté de Médecine, Institut de Chimie Biologique, 11 Rue Humann, 67085 Strasbourg, France

Received 12 June 1981

ABSTRACT.

We have cloned into plasmid pBR322 a DNA fragment extending from position -32 to position -12 of the adenovirus type 2 major late promoter region (position +1 referring to the cap site). *In vitro* transcription experiments show that this 21 base pair sequence, which contains the Goldberg-Hogness or "TATA" box, is both necessary and sufficient for specific initiation of transcription by RNA polymerase B (or II). Furthermore, we show that similar sequences, randomly occuring in the bacterial plasmid pBR322, are also recognized by the RNA polymerase B transcription machinery and able to promote specific *in vitro* transcription. Finally, we discuss the possible importance of the nucleotide sequence of the start region in the actual efficiency of initiation of *in vitro* transcription.

INTRODUCTION.

Elucidation of the mechanism of RNA chain initiation by RNA polymerase B (II) is a prerequisite to understanding the molecular basis of the control of expression of eukaryotic gene coding for proteins. For prokaryotic RNA polymerase some details of this mechanism are known. Briefly stated, the RNA polymerase recognizes and binds to two regions of DNA located about 10 bp and about 35 bp upstream from the start-point of mRNA transcription. These sequences have been termed the promoter (1, 2). The RNA polymerase then locally denatures the DNA around the start-point, selects the first and second ribonucleoside triphosphate and catalyzes the formation of a phosphodiester bond. In bacteria, two forms of control of this mechanism have been demonstrated at the molecular as well as at the physiological level. First, primary sequence variation in the RNA polymerase binding sites leads to different efficiencies of transcription.

Secondly, sequences within and adjacent to the promoter site can serve as binding sites for other regulatory proteins (3). Based on examples of bacterial genetic regulation, control of eukaryotic gene expression at the level of the initiation of transcription has been suggested. Although experimental evidence supporting this idea exists, no clues to its detailed molecular basis have been yet obtained.

We have recently localized a DNA sequence upstream from the 5'-end of the adenovirus-2 major late transcription unit which is necessary for the initiation of *in vitro* specific transcription, by RNA polymerase B. Using *in vitro* genetic techniques, deletions were constructed around the 5'-end of this transcription unit. Deletion of DNA 5' to position -32 and, independently, 3' to position -12 did not abolish specific *in vitro* initiation of transcription (4). In this paper we describe the cloning and transcription analysis of a DNA fragment containing only the -32 to -12 region of the adenovirus-2 major late promoter (Ad2-MLP). Our results show that this 21 base pair sequence, containing the Goldberg-Hogness or "TATA" box, is both necessary and sufficient for specific *in vitro* initiation of transcription by RNA polymerase B. Furthermore, we show that similar sequences, present in the bacterial plasmid pBR322, can also promote specific initiation of *in vitro* transcription by RNA polymerase B.

MATERIALS AND METHODS.

Chemicals and Enzymes.

Restriction enzymes EcoRI and HindIII were prepared according to procedures described (5, 6). Other restriction enzymes were purchased from Boehringer-Mannheim, Bethesda Research Labs. Inc. or New England Biolabs and all reaction conditions were recommended by the suppliers. S1 nuclease was obtained from Miles Laboratories Ltd. (England) and Mung bean nuclease from P.L. Biochemicals Inc.. RNA polymerase B (fraction PCI) was purified from calf-thymus (7). T4 DNA ligase was prepared as described in (8) ; DNA polymerase I was purchased from New England Biolabs and T4 polynucleotide kinase from Boehringer-Mannheim. [α-^{32}P]-CTP and GTP were obtained from

Amersham and [γ-^{32}P]-ATP from New England Nuclear.

Purification of restriction fragments.

Restriction fragments for sequencing, cloning, single-strand-nuclease mapping, and *in vitro* transcription were purified from agarose or polyacrylamide gels or from 5-20 % sucrose gradients containing 10 mM Tris-HCl pH 7.9, 2 mM EDTA, 0.2 M NaCl, 0.1 % Sarkosyl and 5 µg/ml ethidium bromide. The 43 bp ThaI fragment (see Fig. 1A, bottom line) was purified from a 12 % polyacrylamide gel (9) using a DEAE-paper elution technique (10). Single-stranded DNA fragments, labelled with [γ-^{32}P]-ATP on the 5'-end, were purified after denaturation by electrophoresis (4°C, 250 V) on 10 % acrylamide-bisacrylamide (60 : 1) gels, eluted from crushed gel bands and cleaned on DEAE-columns or by filtration through siliconized glass-wool (11).

Construction of clones.

Cloning procedures, preparation of vectors, ligation, transfection in E.coli C600 and screening of colonies were as described (4). Growth of bacteria and extraction of plasmid DNA were essentially as described by Clewell & Helinski (12). Plasmid DNA was purified on ethidium bromide-CsCl equilibrium gradients.

The initial clone was constructed by inserting the BalI fragment E of adenovirus-2 DNA (map position 14.7 - 21.5 on the Ad-2 genome) into the DNA polymerase I repaired EcoRI site of pBR322 ; this clone, called pMLA, contains the complete major-late promoter (Ad2-MLP) region of adenovirus-2 DNA (13). The -32 deletion mutant (Fig. 1A, bottom line) was constructed using the linearized pMLA DNA cut with XhoI (position -259 from the major late RNA startpoint) and treating it with exonuclease III (BRL) and S1-nuclease, to create a series of deletion mutants upstream the "TATA" sequence (4). We selected colonies that had lost the HpaII site (position -51) but retained the PvuII site (position +32) (4). The sequence which replaces the region between the XhoI site and the "TATA" box comes from the 3' end of the adenovirus IVa2 gene which is immediately 5' to the Ad2-MLP and has opposite polarity (14).

The plasmid B and C described in Fig. 1, (B and C respecti-

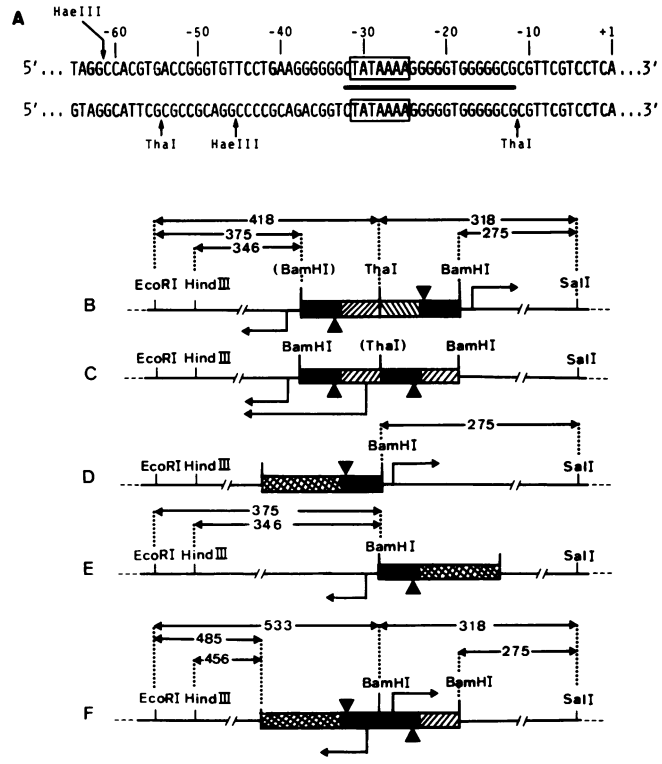

Fig. 1 : Plasmids containing the -32/-12 segment of the Adenovirus 2 major late promoter region.

A) DNA sequence (*non-coding* strand) of the Ad2-MLP region [(4),top line] and of the -32 deletion mutant previously described [(4), bottom line].The solid line between the two sequences indicates the 21 bp of Ad2-MLP DNA present in the plasmids described below. ThaI sites define the -54 to -12 fragment (bottom line), cloned as described in Materials and Methods. Vertical arrows indicate the restriction nuclease cut points. The "TATA" box is boxed. + and - positions are given with respect to the wild type mRNA startsite (position +1).B) plasmid containing two ThaI fragments in opposite orientations (tail to tail, where tail is position -12 and head position -54). C) Plasmid containing two ThaI fragments in the same orientation (head to tail towards the EcoRI site of pBR322). D) Plasmid containing the -61/-12 fragment (A, upper line) of the Ad2-MLP inserted in the BamHI site and orientated towards the SalI site of pBR322). E) Plasmid containing the -61/-12 fragment of the Ad2-MLP in the BamHI site and orientated towards the EcoRI site of pBR322. F) Plasmid containing the ThaI -54/-12 fragment cloned head to head, with respect to the -61/-12 fragment, into the BamHI site of the plasmid described in D. The RNA start sites and the direction of transcription (5' to 3') are indicated by vertical bars followed by horizontal arrows. Lightly shaded areas represent the cloned ThaI -54/-12 fragment and heavily shaded areas the -61/-12 fragment. The underlined sequence in part A is shown in black and the closed triangles indicate the position of the "TATA" box. Restriction sites in parentheses have been lost during cloning. For details of cloning procedures see Materials and Methods. Numbers between horizontal arrows correspond to distances in base pairs (17).

vely) were constructed by inserting the 43 bp ThaI fragment purified from the -32 mutant (Fig. 1A, bottom line) into pBR322. This fragment contains the sequences from the ThaI site at -12 (between the start-point and the "TATA" box) and extends 43 bp in the 5' direction to a ThaI site (position -54) in the replacement sequence of the -32 deletion mutant cross-hatched box in Fig. 1B and C. This fragment, which contains only the sequence from -32 to -12 of the Ad2-MLP (heavy line in Fig. 1A and black box in Fig. 1, B to F) was cloned into the DNA polymerase I repaired BamHI site of pBR322 DNA. Colonies were picked at random and grown as small cultures. Colonies containing the ThaI fragment were detected by analysis of clear lysate DNA (15) with restriction enzyme HaeIII. Digestion products were analyzed on 10 % polyacrylamide gels. The structure of several clones was deduced from the regeneration of restriction sites and confirmed by sequencing (11). Because a 5-fold molar excess of the ThaI fragment was used in this cloning procedure, all colonies characterized contain tandem insertions of the fragment. Clone B (Fig. 1B) has two fragments in opposite orientation (tail to tail), while clone C has two fragments in the same orientation (toward the EcoRI site of pBR322, head to tail).

Plasmids D and E (Fig. 1) were constructed as previously described (4). Briefly, a 91 bp HaeIII restriction fragment, from -61 to +30 of the Ad2-MLP (Fig. 1A, topline), was inserted in the DNA polymerase I repaired EcoRI site of pBR322. From a complete digestion of this plasmid with restriction enzyme ThaI, a 158 bp fragment was purified, which contains the -61/-12 region of the Ad2-MLP (black + crossed box in Fig. 1D and E) and, upstream, a piece of pBR322 (not shown as a box in Fig. 1). This fragment was subcloned in both orientations in the repaired BamHI site of pBR322. After cloning only the BamHI site closest to the "TATA" sequence was restored in clone D.

Plasmid F was constructed by inserting the purified 43 bp ThaI fragment into the unique BamHI site of plasmid D after making it blunt by repair with DNA polymerase I. Screening for colonies containing the inserted fragments was carried out as for clones B and C using HaeIII digestion and analysis by polyacrylamide gel electrophoresis. Plasmid F contains a

single insertion of the 43 bp ThaI fragment in a head to head orientation with respect to the -61/-12 region contained in plasmid D. The DNA sequence of this clone was inferred from the regeneration of the BamHI sites.

Specific *in vitro* transcription.

RNA was synthesized *in vitro* for 60 minutes at 25°C in a standard reaction mixture (final volume 25 µl) containing (final concentration) 25 µM [α-^{32}P]-CTP or GTP (8000 counts/min/pmol), 500 µM ATP, 500 µM UTP and 500 µM of GTP or CTP, 10 mM Tris-HCl pH 7.9, 7.5 mM $MgCl_2$, 50 µM KCl, 10 % glycerol, 0.25 mM dithiotreitol, 0.120 units of calf-thymus RNA polymerase B (7) 12.5 µl of HeLa cell S100 extract (16) and 0.5-1.0 µg, of template DNA. RNA synthesis was stopped with 450 µl of a solution containing 500 mM Na-acetate pH 5.0, 0.5 % SDS, 50 µg/ml of E.Coli tRNA ; RNA was extracted with water saturated phenol, twice with chloroform, made 0.15 M in ammonium acetate and precipitated with two volumes of ethanol. The RNA pellet was dissolved in 20 µl formamide buffer (11), denatured 5 min. at 65°C, chilled on ice, and analyzed on a 3.5 % or 5 % acrylamide -8.3 M urea gel (11).

Single-strand nuclease mapping.

The localization by single-strand nuclease mapping of the 5' end of RNAs synthesized *in vitro* was made essentially as described previously with minor modifications (4, 13). RNA was synthesized *in vitro* (three-fold standard reactions, one aliquot was used to check the synthesis on a transcription gel) on the SalI linearized -32/-12 and -61/-12 plasmids B and D, extracted, precipitated with 0.5 pmole of single-stranded BamHI-Sau3A probe [5'-end-labelled with ^{32}P at the Sau3A site (17) with a specific activity of 500.000 cpm/pmol of 5'-end], resuspended and hybridized for 16 hours at 42°C. The S1-nuclease reaction was performed with 300 units of S1-nuclease (Miles) at 25°C and stopped after 60 min. The Mung bean nuclease reaction was performed at 25°C for 40 minutes with 500 units of Mung bean nuclease in a mixture (250 µl) containing 0.03 M sodium acetate (pH 4.6), 0.05 M NaCl, 1 mM $ZnCl_2$, 5 % glycerol and 15 µg/ml of sonicated and denaturated calf-thymus DNA.

The reactions were stopped with 50 mM EDTA (final concentration) and 20 µg of carrier E.coli tRNA and, after phenol extraction made 0.15 M in ammonium acetate and precipitated with two volumes of cold ethanol. Products resistant to S1 or Mung bean nuclease digestions were analyzed on 8 % or 10 % polyacrylamide, 8.3 M urea sequencing gels (11).

RESULTS AND DISCUSSION.

A) The "TATA" region is necessary and sufficient to direct specific initiation of RNA synthesis.

1. Promotion of transcription initiation by the -32/-12 segment of the Adenovirus-2 major-late promoter.

To determine whether the 21-bp region from -32 to -12 of the Ad2-MLP is sufficient to direct specific initiation of transcription *in vitro*, we have cloned in pBR322 a 43-bp fragment which contains this region. Lines B and C of Figure 1 represent two plasmids which harbour in this BamHI site two such fragments in opposite (B) or identical (C) orientations (see Materials and Methods for cloning details). Lines D and E show the maps of the previously described (4) -61/-12 fragment of Ad2-MLP cloned in the BamHI site of pBR322. Line F shows the map of a plasmid which contains in opposite orientation both the -32/-12 and -61/-12 regions of the Ad2-MLP.

When plasmid B is linearized with restriction enzymes EcoRI, HindIII or SalI and transcribed *in vitro*, RNA run-offs of about 365 (Fig. 2 lane 2), 335 (lane 3) and 265 (lane 4) nucleotides are obtained, respectively. At optimal DNA concentrations the amount of specific run-off transcripts for the -32/-12 templates is about 10 % of the wild-type (lane 1). Comparing run-off transcripts from the -32/-12 plasmids B to those from the -61/-12 plasmid E and D templates cut with the same restriction enzyme (Figure 2, compare 2, 3 and 4 to 7, 8 and 9, respectively) we see that the length as well as the intensity of the specific transcripts is about the same. When the template B is cut with both SalI and EcoRI (lane 5) or HindIII (lane 6) both run-offs are observed. Because so few template molecules are apparently utilized in this *in vitro*

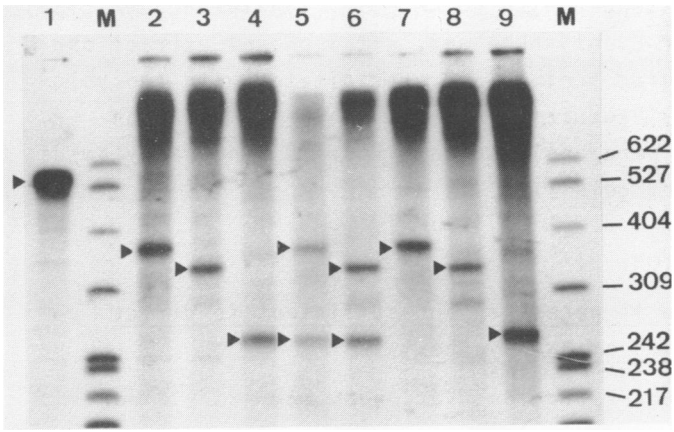

Fig. 2 :Comparison of specific *in vitro* transcription from -32/-12 (B in Fig. 1) and -61/-12 (D and E in Fig. 1) plasmids which contain fragments of the Ad2-MLP. DNA was linearized with single-cut restriction enzymes and used as template to generate specific run-off transcripts (Material and Methods). After synthesis the reaction mixture was processed as described in Materials and Methods and RNA was analysed on a 5 % acrylamide -8.3 M urea gel. Lane 1, purified TaqI fragment (positions -250/+550 on the Ad2-MLP) from wild-type Ad2-MLP plasmid DNA (13) ; lane 2, plasmid B (Fig. 1) cut with EcoRI ; lane 3, plasmid B cut with HindIII ; lane 4, plasmid B cut with SalI ; lane 5, plasmid B cut with EcoRI + SalI ; lane 6, plasmid B cut with HindIII + SalI ; lane 7, plasmid E cut with EcoRI ; lane 8, plasmid E cut with HindIII ; lane 9, plasmid D cut with SalI. Lanes M are 5'-end ^{32}P- labelled DNA fragments (pBR322 plasmid DNA cut with MspI) used as size markers. Arrowheads point to the run-off transcripts discussed in the text.

transcription system (16), we can say nothing about the possible interference between two promoters when in close proximity. Transcription of plasmids C and F (Fig. 1) is shown in Fig. 3. With plasmid C, two run-offs in the EcoRI direction are observed on templates cut with either EcoRI (lane 2, about 410 and 365 nucleotides) or HindIII (lane 3, about 380 and 335 nucleotides). When plasmid F, cut with EcoRI, HindIII or SalI is used as template, the transcripts (about 515, 485 and 300 nucleotides, respectively) shown in lanes 4-6 are obtained. Inspection of lanes 7 and 8 reveals no bands of similar sizes when wild-type pBR322 digested with EcoRI or SalI is transcribed and the products are electrophoresed under the same conditions.

Several conclusions can be drawn from these transcription data. First, from a length analysis of all of the run-off transcripts shown in Fig. 2 and 3 it is clear that in all cases the "TATA" region directs RNA polymerase to initiate specifically

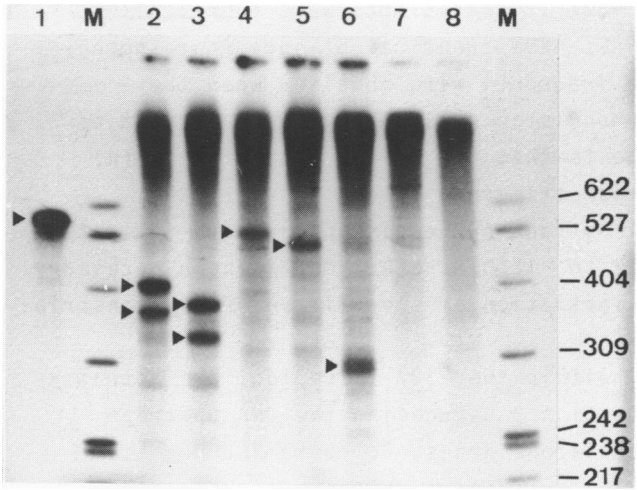

Fig. 3 : Analysis of *in vitro* specific RNA transcripts synthesized on linearized plasmids C and F (see Fig. 1). RNA synthesis and analysis were as in Fig. 2. DNA templates were as follows : lane 1, purified TaqI fragment (position -250/+550 on the Ad2-MLP) from wild type Ad2-MLP plasmid DNA (13) ; lane 2, plasmid C cut with EcoRI ; lane 3, plasmid C cut with HindIII ; lane 4, plasmid F cut with EcoRI ; lane 5, plasmid F cut with HindIII ; lane 6, plasmid F cut with SalI ; lane 7, pBR322 wild-type plasmid cut with EcoRI ; lane 8, pBR322 wild-type plasmid cut with SalI. Size markers (M) and arrowheads are as in Fig. 2.

about 30 bp downstream from the first T of the "TATA" box. Second, deletion of the -61/-33 segment from the -61/-12 region does not change the apparent start point of *in vitro* transcription within the limits of resolution of the gel system employed (4 bp). Third, a comparison of all of the run-off transcripts with the transcript obtained with a template containing the entire major late promoter region (lane 1 in Fig. 2 and 3) indicates that the amount of specific transcription from the -32/-12 and the -61/-12 plasmids is 5 to 10-fold reduced ; in addition deletion of the -61/-33 segment from the -61/-12 region does not reduce the initiation of specific transcription below that promoted by the -61/-12 region itself. Fourth, the observation of both specific run-off transcripts with plasmid C (Fig. 1 and Fig. 3 lanes 2 and 3) excludes that the "TATA" sequence could serve as a processing signal, as only the shorter run-off would be expected.

To determine more precisely the start point of the run-off transcripts directed by the -32/-12 region, we used the same S1-mapping strategy as previously used to map the start points of the -61/-12 transcripts (4). The probe DNA fragments protected

by the -61/-12 and the -32/-12 transcripts were identical, indicating that these RNAs have identical 5'-ends (not shown). Moreover the same probe fragments were obtained when the single-strand specific Mung bean nuclease was used instead of S1 nuclease, which suggests that the 5'-end multiplicity (4) was not due to S1 nuclease trimming. Taken together, the transcription and single-strand nuclease mapping data indicate that the region from -32 to -12 of the Ad2-MLP is both necessary and sufficient for the initiation of specific *in vitro* transcription.

Which nucleotides, within the -32/-12 region, are actually involved in promoter function ? Sequencing the DNA upstream from several Drosophila histone genes, Goldberg and Hogness (personal communication) identified a sequence of homology, the "TATA" box, which since has been found about 30 bp upstream from the cap site of most genes transcribed by RNA polymerase B (4, 18). Analysis of these 5' upstream sequences for 60 genes yields the consensus sequence 5'-TATA$^{A}_{T}$A$^{T}_{A}$-3' (4, 18). Corden et al. (4) have shown that deletion of the first TA of the Ad2-MLP TATAAA sequence reduces specific *in vitro* transcription by at least 100-fold. Deletion of the entire "TATA" box eliminates specific *in vitro* transcription from this promoter. Furthermore, Wasylyk et al. have shown that some point mutations and deletions in the TATAAA sequence reduces the specific *in vitro* transcription of the chicken conalbumin gene by about 20-fold (19, 20). The region adjacent to the "TATA" box homology sequence shows little or no conservation among the 60 genes which have been examined (18 and J. Corden, unpublished). In considering the lack of homology other than the "TATA" box in the -32/-12 region and the effect of deletions and point mutations in this sequence, we are led to hypothesize that the "TATA" box is the sequence element which is necessary and sufficient for the initiation of specific *in vitro* transcription.

2. Promotion of specific initiation of transcription by pBR322 "TATA" box-like sequences.

To test the above hypothesis, we located, through a computer search, sequences in pBR322 which are homologous to the eucaryotic "TATA" sequence. Two sequences are of particular

interest : the 11 base sequence 5'-GGACTATAAAG-3' (Fig. 4A) starting at nucleotide 2608 (21) shares 11, 10 and 7 nucleotides with the "TATA" box region of the rat preproinsulin gene (22), the human insulin (23), and the Ad2-MLP, respectively ; at position 4320 (Fig. 4A) we find the 8 base sequence 5'-CTATAAAA-3' which corresponds exactly to 8 nucleotides of the Ad2-MLP and chicken conalbumin (24) "TATA" box region. Both of these sequences are located at positions quite different from the bacterial promoters used in the pBR322 plasmid (25). When wild type pBR322 DNA is linearized with the appropriate restriction enzymes, transcribed *in vitro*, and the run-off products are analyzed on gels,we find transcription run-off products which correspond to initiations directed by these bacterial "TATA" box-like sequences. Fig. 4 shows examples of such *in vitro* run-off transcripts, using as template pBR322 DNA linearized at different restriction sites. With templates linearized with PvuI, PstI and BglI (see Fig. 4, A), transcripts of respectively about 1100 (Fig. 4B, lane 1), 970 and 840 nucleotides (not shown) are obtained. When the purified 434-bp HaeIII fragment (mapping between position 2518 and 2952 of pBR322, see Fig. 4, A) is used as template a transcript of about 310 nucleotides is observed (Fig. 4C, arrowhead). Together the lenghts of these run-off transcripts allow us to position their startpoint approximately 30 bp downstream from the bacterial "TATA" box-like sequence located at position 2608. Similarly, the lenghts of the run-offs obtained on pBR322 templates cut with SalI (655 nucleotides, Fig. 4B, lane 3) or BamHI (385 nucleotides, not shown) extrapolate to a start site located about 35 bp downstream from the "TATA" box-like sequence found at position 4320. When the pBR322 templates are transcribed in the presence of 1 µg/ml α-amanitin (Fig. 4B, lanes 2 and 4) the specific run-off bands are not seen, indicating that these transcripts are synthesized by RNA polymeraseB. This specific initiation of transcription on a bacterial template supports the conclusion that the "TATA" sequence alone is sufficient for promoting specific *in vitro* transcription. If an additional sequence is required for specific initiation, it would seem unlikely to occur in the pBR322 genome adjacent to these "TATA" box-like sequences. Perhaps transcription from

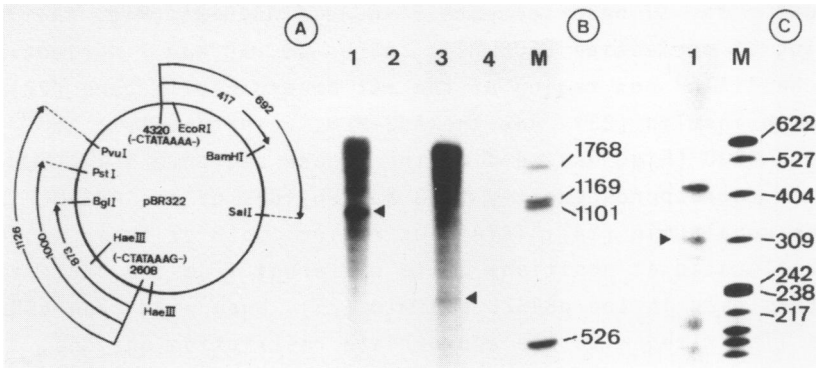

Fig. 4 : Specific *in vitro* transcription of pBR322 plasmid DNA. A) Physical map of pBR322 DNA showing the location of the "TATA" box-like sequences (2608 and 4320 refer to the position of the first T present in the sequences given in parentheses) and of restriction sites used to generate the DNA templates (17). The lengths of the various DNA fragments covered by arrows are given in base pairs. The distances from position 2608 to the two HaeIII sites, shown on the diagram, are 344 and 90 base pairs in the clock-wise and counter clock-wise directions respectively. Other "TATA"-like sequences have been localized in pBR322 by computer search but have not been further investigated : these sequences are located at positions 97, 443, 3233, 3271, 3947 and correspond respectively to the "TATA" box sequences of the genes coding for histone H4 (43), H1 (44), lysozyme (Grez, M., Schütz, G., Jung, A. and Sippel, A., personnal communication), ovalbumin (38) and lysozyme (Grez, M., Schütz, G., Jung, A. and Sippel, A., personal communication). B) pBR322 DNA was linearized with different restriction enzymes and used as template in the *in vitro* transcription system described in Materials and Methods. The RNA transcripts were analysed on a 3.5 % acrylamide 8.3 M urea gel. Lane 1, the pBR322 template was cut with PvuI (position 3737) ; lane 2, same as in lane 1, but α-amanitin (1 µg/ml) was present during RNA synthesis ; lane 3, pBR322 was cut with SalI (position 650) ; lane 4, same as in lane 3, but α-amanitin was present (1 µg/ml) during RNA synthesis ; lane M is 5'-end ^{32}P labelled DNA fragments (SV40 DNA cut with HindIII) used as size markers. Lanes 3 and 4 have been exposed about twice as long as lanes 1 and 2. C) the 434 bp HaeIII restriction fragment (see part A) was electrophoretically purified and used as template for *in vitro* RNA synthesis. RNA transcripts were analysed on a 5 % acrylamide 8.3 M urea gel (lane 1). The upper band seen on lane 1 corresponds to full length template DNA which becomes end labelled during the *in vitro* incubation (J. Corden, unpublished observation). Lane M is 5'-end 32P-labelled DNA fragments (pBR322) cut with MspI used as size markers.

these sequences of pBR322 or from similar sequences in other plasmids could account some of the spurious transcription observed when recombinant plasmids are injected into Xenopus oocytes (26, 27).

B) The sequence of the startsite influences the efficiency of initiation of *in vitro* transcription.

We have previously shown that sequences downstream from position -12 are important for the efficiency of specific initiation of *in vitro* transcription on both the adenovirus-2 and chicken conalbumin genes. Replacing the startsite sequences of these genes with a plasmid sequence resulted in a 10-fold reduction in transcription efficiency. Here we show that when the same plasmid sequence replaces the startsite sequence of the -32 deletion of the Ad2-MLP the same reduction in efficiency is observed (Fig. 2 and 3). Analysis of sequences around the startpoints of 22 eucaryotic genes where the 5'-end of the mRNA has been accurately mapped indicates a preferred startpoint sequence consisting of an A residue, 29 to 33 nucleotides downstream from the 5'-end of the "TATA" box sequence, surrounded by pyrimidines (4, 18). The apparent preference for initiation on A residues is consistent with the observation that 7mG pppAp is the most common 5' mRNA "cap" structure found in eucaryotic mRNA (28). An examination of the startpoint sequences of the various plasmids described in this paper which contain the "TATA" box of the Ad2-MLP (Fig. 5, lines 1-4) shows the presence of an A residue within 4 nucleotides of the corresponding wild-type +1 position. Since all of these clones behave as down mutants it is clear that some other sequence requirement exists. For example clone F in the EcoRI direction (line 4) contains the exact consensus startsite sequence (i.e. an A residue surrounded by runs of pyrimidines) yet it is less efficiently transcribed than wild-type. Perhaps the high GC content of this startsite region makes the starting nucleotide less accessible. Analysis of the sequence downstream from the bacterial "TATA" box-like sequences (Fig. 5) shows that the predicted 4320 startpoint region (line 5) contains no A residue within the 10 bases surrounding the predicted +1 position (the closest A is 37 nucleotides from the 5'-end of the "TATA"

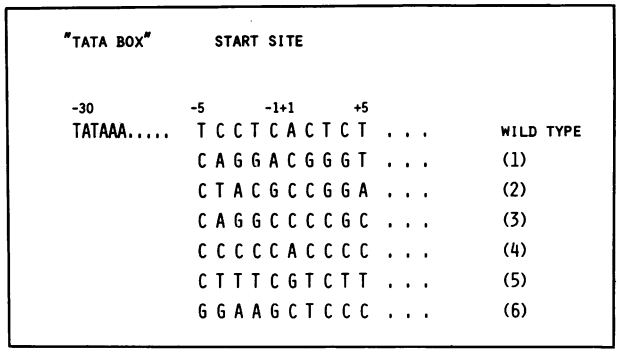

Fig. 5 : Nucleotide sequences around the RNA startpoints of the wild-type Ad2-MLP, the different Ad2-MLP containing plasmids and the "TATA" box-like sequences of pBR322.
Ten bases of the DNA sequence (*non coding* strand) in the RNA startpoint region equidistant from the "TATA" sequence, are represented for the different plasmid DNAs (1 to 6) below the corresponding Ad2-MLP wild-type sequence (top line). Startpoint regions (see Fig. 1) : (1) in plasmids B and E with transcription occuring in the EcoRI direction ; (2) plasmids B and D in the SalI direction ; (3) plasmid C with transcription directed by the "TATA" sequence closest to the SalI site ; (4) plasmid F with transcription occuring in both directions, (5) and (6) pBR322 with transcription directed by the "TATA" box-like sequences, located at position 4320 and 2608, respectively.

sequence). The predicted 2608 startpoint has A's in position -2 and -3 (line 6). The intensity of the specific run-off transcript from the 2608 "start" is about the same as for the plasmids containing the Ad2-MLP -32/-12 or -61/-12 promoter sequences, but the intensity of the 4320 start is at least 10-fold lower. The inefficiency of the 4320 start might then be related to the distance of an A residue at about 30 bp downstream from the pBR322 "TATA" like sequence.

CONCLUSION

The present results demonstrate that the Adenovirus major late "TATA" box region is able to promote *in vitro* specific transcription from sites located at about 30 bp downstream, even when separated from both its natural upstream and downstream sequences and inserted into a foreign nucleotide environment. Moreover randomly occuring prokaryotic "TATA" box-like sequences

are recognized by the eukaryotic RNA polymerase B transcription machinery and promote specific *in vitro* transcription. Several recent studies support the conclusion that the Adenovirus-2 major late and chicken conalbumin "TATA" box sequences are essential for specific *in vitro* transcription (4, 13). Mathis and Chambon (29) have also shown that deletion of sequences upstream from the SV40 early "TATA" box do not abolish specific *in vitro* transcription, whereas deletion of the "TATA" box does. Similar deletion experiments by Tsai et al. (30) on the chicken ovalbumin gene indicate that the "TATA" box plays an essential role in promoting specific *in vitro* transcription. Results of transcription of "TATA" box point mutants further supports this contention. Wasylyk et al. (19) have shown that T to G or T to A (20) transversions resulting in "TAGA" or "TAAA" sequences upstream to the chicken conalbumin gene act as 20-fold promoter down mutations. A similar sequence change (T to G transversion) occurs in the second T of the "TATA" box of one of the mouse α-globin pseudogenes (31). This gene is not transcribed *in vitro* or *in vivo*, although other sequence changes, particularly around the cap site, may contribute to this effect (31).

Although our results indicate that the *in vivo* cap site sequences are not essential for obtaining specific discrete transcripts *in vitro*, it is clear that when these sequences are replaced the efficiency of transcription is reduced. We have previously noted that cap-site sequences consist very often of an A residue within a pyrimidine-rich cluster (4, 18). The results presented here support this general sequence requirement although many more start-site replacement mutants will need to be analyzed before specific sequence requirements can be determined.

While removal of sequences downstream from the "TATA" box can reduce the efficiency of *in vitro* transcription, removal of upstream sequences has no *in vitro* effect (4, 13, 29). In contrast, when promoters with deletions upstream from the "TATA" box are analyzed *in vivo*, sequences in the -150 to -60 region play an important role. For the SV40 early (32, 33), Herpes thymidine kinase (34, and K. O'Hare personal communication), yeast His3 (35) and histone H_2A (36, 37) promoters, deletion of

sequences upstream to the "TATA" box results in a reduction in specific transcription. Sequence homology between different promoters has been observed in this "upstream" region ("CAAT" box, see Ref. 38), although whether this sequence itself actually contribute to the promotion of transcription has not yet been determined. In view of the present low efficiency of *in vitro* transcription, it is quite possible that the *in vitro* system used to date lacks some essential factor(s) necessary for recognition of the entire *in vivo* promoter region.

Whatever the role of these upstream sequences is it seems clear that the "TATA" box is necessary both *in vitro* and *in vivo* for transcription to start at the *in vivo* cap sites. It has been shown that deletion of the "TATA" box does not abolish *in vivo* gene expression (32, 36). However, when the "TATA" box sequences are deleted from the SV40 early (33), histone H_2A (37), and rabbit β-globin (39) promoters the natural mRNA's starts are not seen, but other new multiple starts appear and these starts are the same as those used *in vitro* in the absence of the TATA sequence (29). These results and the absence of the TATA sequence in the promoter region of the adenovirus EII and IVa2 genes and of the papovavirus late genes (40, 41, 42) indicates also that some other types of RNA polymerase B promoter sequences should exist. Further understanding of the role of various eucaryotic promoter sequences should come shortly from a detailed biochemical analysis of the factor(s) required for *in vitro* transcription as well as from continued analysis of promoter sequence modifications *in vivo*.

ACKNOWLEDGEMENTS.

We thank J.M. Jeltsch for helpful advice during our sequence work, J.M. Garnier for generous gifts of pBR322 plasmid DNA, K. Dott and D. Schmitt for growing cells and virus, E. Badzinski and B. Boulay for help during preparation of the manuscript. We are particularly grateful to C. Hauss for continuous technical assistance. This investigation was supported by grants from the INSERM (PRC de Génie Génétique), from the CNRS (ATP Biologie Moléculaire du Gène, contract n°. 006520/60), from the "Association pour le Développement de la Recherche sur le Cancer", from

the Fondation pour la Recherche Médicale Française and from the Fondation Simone et Cino del Duca.

Abbreviations :

bp, base pairs ; Ad2-MLP, Adenovirus2-Major Late Promoter ; SV40, Simian Virus 40.

REFERENCES.

1. Pribnow, D. (1975) Proc. Natl. Acad. Sci. USA 72, 784-788.
2. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
3. Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) Cell 20, 269-281.
4. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kédinger, C. and Chambon, P. (1980) Science 209, 1406-1414.
5. Sümegi, J., Breedveld, D., Hossenlopp, P. and Chambon, P. (1977) Biochem. Biophys. Res. Commun. 76, 78-85.
6. Humphries, P., Gordon, R.L., McConnell, D.J. and Connolly, P. (1974) Virology 58, 25-31.
7. Kédinger, C. and Chambon, P. (1972) Eur. J. Biochem. 28, 283-290.
8. Tait, R.C., Rodriguez, R.L. and West, R.W. (1980) J. Biol. Chem. 255, 813-815.
9. Maniatis, T., Jeffrey, A. and Van de Sande, H. (1975) Biochemistry 14, 3787-3794.
10. Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) Anal. Biochem., in press.
11. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA, 74, 560-564.
12. Clewell, D.B. and Helinski, D.R. (1969) Proc. Natl. Acad. Sci. USA, 62, 1159-1166.
13. Wasylyk, B., Kédinger, C., Corden, J., Brison, O. and Chambon, P. (1980) Nature 285, 388-390.
14. Akusjärvi, G. and Pettersson, U. (1979) J. Mol. Biol. 134, 143-158.
15. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
16. Weil, P.A., Luse, D.S., Segall, J. and Roeder, R.G. (1979) Cell 18, 469-484.
17. Sutcliffe, J.G. (1978) Nucleic Acids Res. 5, 2721-2728.
18. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, in press.
19. Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Téoule, R. and Chambon, P. (1980) Proc. Natl. Acad. Sci. USA, 77, 7024-7028.
20. Wasylyk, B. and Chambon, P. (1981) Nucleic Acids Res. 9, 1813-1824.
21. Sutcliffe, J.G. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 77-90.
22. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979) Cell 18, 545-558.

23. Bell, G.I., Pictet, R.L., Rutter, W.J., Cordell, B., Tischer, E. and Goodman, H.M. (1980) Nature 284, 26-32.
24. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. and Chambon, P. (1979) Nature 282, 567-574.
25. Stüber, D. and Bujard, H. (1981) Proc. Natl. Acad. Sci. USA 78, 167-171.
26. Wickens, M.P., Woo, S., O'Malley, B.W. and Gurdon, I. (1980) Nature 285, 628-634.
27. Trendelenburg, M.F., Mathis, D., Oudet, P. (1980) Proc. Natl. Acad. Sci. USA 77, 5984-5988.
28. Lewin, B. (1980) Gene Expression, Eucaryotic Chromosomes, 2nd edn. pp. 668-675, John WILEY & Sons, Inc.
29. Mathis, D.J. and Chambon, P. (1981) Nature 290, 310-315.
30. Tsai, S.Y., Tsai, M.J. and O'Malley, B.W. (1981) Proc. Natl. Acad. Sci. USA 78, 879-883.
31. Talkington, C.A., Nishioka, Y. and Leder, P. (1980) Proc. Natl. Acad. Sci. USA 77, 7132-7136.
32. Benoist, C. and Chambon, P. (1980) Proc. Natl. Acad. Sci. USA 77, 3865-3869.
33. Benoist, C. and Chambon, P. (1981) Nature 290, 304-310.
34. Mc Knight, S.L. and Gavis, E.R. (1980) Nucleic Acids Res. 8, 5931-5948.
35. Struhl, K.(1981) Proc. Natl. Acad. Sci. USA, in press.
36. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 1432-1436.
37. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 7102-7106.
38. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142.
39. Dierks, P., Van Ooyen, A., Mantei, N. and Weissmann, C. (1981) Proc. Natl. Acad. Sci. USA 78, 1411-1415.
40. Baker, C.C., Herisse, J., Courtois, G., Galibert, F. and Ziff, E. (1979) Cell 18, 569-580.
41. Baker, C.C. and Ziff, E. (1981) J. Mol. Biol., in press.
42. Flavell, A.J., Cowie, A., Arrand, J.R. and Kamen, R. (1980), J. Virology 33, 902-908.
43. Busslinger, M., Portmann, R., Irminger, J.C. and Birnstiel, M.L. (1980) Nucleic Acids Res. 8, 957-977.
44. Sures, I., Levy, S. and Kedes, L.H. (1980) Proc. Natl. Acad. Sci. USA 77, 1265-1269.