Expression of Cloned Genes in New Environment

Promoter Sequences of Eukaryotic Protein-Coding Genes

J. Corden, B. Wasylyk, A. Buchwalder, P. Sassone-Corsi
C. Kedinger, P. Chambon

Summary. In vitro genetic techniques were used to study the sequence requirements for the initiation of specific transcription. Deletion mutants were constructed around the putative promoter of the adenovirus-2 major late and chicken conalbumin genes. Specific transcription in vitro by RNA polymerase B together with a HeLa cell cytoplasmic extract was used as the test for promoter function. With this approach sequences which are essential for the initiation of specific transcription in vitro, were shown to be located between 12 and 32 base pairs upstream from the 5′ end of these genes.

A basic property of all living cells is the ability to switch the expression of their genes on and off, for example, in response to extracellular signals. In prokaryotes this switching is mostly controlled at the level of RNA transcription. In complex eukaryotic organisms, although the expression of many different genes is turned on and off during development from the egg, and although this switching continues in the differentiated cells, the importance of control of gene expression at the transcriptional as opposed to posttranscriptional level is still a matter of controversy. In higher eukaryotes, however, there is unequivocal evidence that the expression of at least some genes is controlled at the level of RNA transcription. For example, in the tubular gland cells of the magnum portion of the chick oviduct, the transcription of the ovalbumin and conalbumin (ovotransferrin) genes is turned on by the steroid hormones estradiol and progesterone. Like the genes coding for the other egg white proteins, these genes are not transcribed in the absence of estradiol or progesterone [see (*1*) and references therein].

During the past 20 years some of the basic mechanisms involved in regulation of transcription in prokaryotes and their viruses have been elucidated in molecular terms. It has been learned that transcription is regulated by modulation of the efficiency with which RNA polymerase can recognize and interact with specific DNA signal sequences (promoters and terminators) that specify starting or stopping sites and are involved in the promotion and termination of RNA transcription [see (*2*) and references therein]. The genetic approach has been invaluable in these studies. For instance, it is primarily on genetic evidence that Jacob *et al.* (*3*) first defined the promoter as an initiating element indispensable for the expression of bacterial structural genes. Further progress was made possible by the availability in vitro of cell-free systems, reconstructed from purified components, in which the selective in vivo transcription events could be accurately duplicated.

In addition to requiring purified RNA polymerase and well-defined templates, such studies in vitro also require a detailed knowledge of the transcription unit in vivo to determine whether correct initiation and termination of transcription are occurring. The use of such in vitro systems, the possibility of purifying specific wild-type and mutant prokaryotic genes and their RNA products, and the availability of DNA and RNA sequencing methods have enabled investigators to analyze the structure and function of certain genes in great detail and to show that prokaryotic promoters are regions of DNA 5′ to the structural genes (*2*, *4*).

The messenger RNA (mRNA) start points (the position on a DNA sequence which codes for the first nucleotide of an RNA) of many prokaryotic transcription units have been precisely located by genetic analysis and transcription in vitro, and the DNA sequences of these regions have been determined. Pribnow (*5*) and Schaller (*6*) first noted a sequence homology, related to 5′-TATAATG-3′ (T, thymine; A, adenine; G, guanine), located about 10 base pairs (bp) upstream from mRNA start points (*7*). A second region of homology, the "recognition region," has also been noted in some promoters in a region centered about 35 bp upstream from the mRNA start point (*2*). Deoxyribonuclease protection and chemical modification experiments have been used to show that RNA polymerase binds to these regions (*2*). Furthermore, several promoter mutants have been sequenced, and their locations within the homologous sequences have established

The authors are staff members at the Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, Strasbourg, France.

0036-8075/80/0919-1406$02.00/0

that these regions fit the original definition of a promoter.

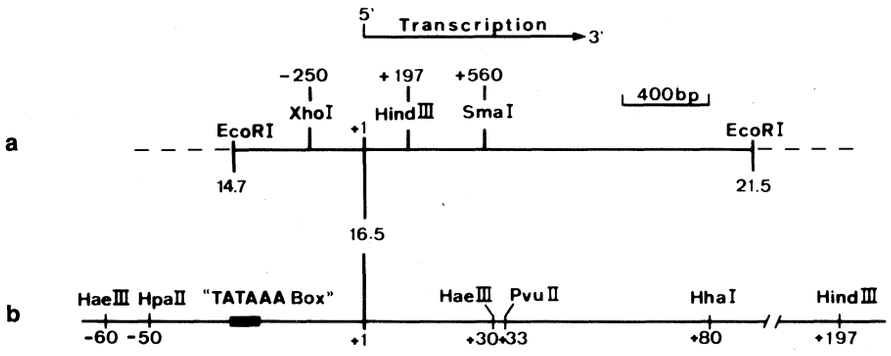
In contrast to prokaryotic cells, the molecular mechanisms that underlie the regulation of transcription in eukaryotic cells are still largely unknown, notably because the classical genetic approach is for the most part not possible in these cells. That the mechanisms in eukaryotes may not be identical to those in prokaryotes was first suggested 10 years ago by the discovery of the multiplicity of eukaryotic RNA polymerases by our group and Roeder and Rutter [for reviews, see (*8, 9*)]. It was subsequently established that cells of both higher and lower eukaryotes contain three structurally and functionally distinct classes of RNA polymerase which are localized in different subcellular fractions. Class A or I catalyzes the synthesis of ribosomal RNA, class B or II that of mRNA, and class C or III that of transfer RNA (tRNA) and 5*S* RNA [for reviews, see (*8, 9*)].

Although highly purified preparations of these enzymes, particularly RNA polymerase B, were available shortly after their discovery, progress has been extremely slow in analyzing their role in the control of transcription. The lack of meaningful cell-free transcription systems in vitro mostly accounts for this failure. Indeed, because of the complexity of the eukaryotic genome, there was no means to study the transcription of a given gene in vitro by incubating the total cellular DNA with purified RNA polymerase. Even when well-defined viral DNA templates such as the Simian virus 40 (SV40) and adenovirus-2 genomes were available, the primary transcription products were unknown, precluding any valid analysis of the factors involved in the control of transcription. Furthermore, intact viral DNA's proved to be very poor templates in vitro for the purified RNA polymerase B, which was known to transcribe SV40 and adenovirus genomes in vivo (*8*). Several technical breakthroughs were clearly required.

The discovery of restriction enzymes and reverse transcriptase, followed by the advent of molecular cloning and of methods for separating, visualizing, and rapidly sequencing DNA and RNA molecules, have made it possible to study, at the nucleotide level, the anatomy of eukaryotic cellular and viral genes and of their primary RNA transcripts (*10, 11*). It has now been shown in several instances [for example, see (*12*) and (*13*) for the adenovirus major late transcription unit and for the ovalbumin transcription unit, respectively] that the 5′ ends of the RNA primary transcripts and of the mature mRNA's coincide and therefore that the start point of transcription corresponds to the base coding for the 5′ terminal nucleotide of the mRNA's. By analogy to the situation in bacteria, we would expect eukaryotic promoters to be located in the region adjacent to the 5′ end of the transcription unit. Indeed, the comparison of several cellular and viral genes has revealed the existence of an AT-rich region of homology centered about 25 bp upstream from the mRNA start points (*14–16*). This sequence, which is known as the "TATA" box, was first noticed by Goldberg and Hogness and bears some sequence resemblance to the Pribnow box in prokaryotic promoters (*14*) (see below).

However, this homologous sequence is not found upstream from the start point of genes transcribed by RNA polymerases A (*17*) and C (*18, 19*), indicating that the specific transcription of different classes of genes by the distinct classes of eukaryotic RNA polymerases could be due to the specific recognition of sequences characteristic of a class of genes.

Although the recognition of sequence homologies is important in suggesting the location of control regions, it is obvious that the actual role of homologous sequences cannot be established without a functional assay, for instance, a cell-free system capable of accurate in vitro transcription. A technical breakthrough



c

```
          -50       -40       -30       -20       -10
          CCGGGTGTTCCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCG

(-21)     CCCCGCAGACGGTCTCGCATTCCACGAGCCAGGGTGGGGGCGCG
(-29)     CCAGTGAGCTCTGGCCGTTCGGGGTAAAAGGGGGTGGGGGCGCG
(-32)     GCCGCAGGCCCCCGCAGACGGTCTATAAAAGGGGGTGGGGGCGCG
(-41)     CAGGCCCCGCAGGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCG
(-43)     GGCCAGTGAGCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCG
(-44)     GAGCTCTGGCCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCG
(-45)     GTCTCGCATCCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCG
(-47)     CCCCGCGTTCCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCG
```

Fig. 1. Construction in vitro of mutants with deletions upstream from the 5′ end of the adenovirus-2 major late transcription unit. The plasmid pMLA was constructed by inserting the Bal I fragment E of adenovirus-2 DNA (map units 14.7 to 21.5, line a) into the Eco RI site of pBR322. The pMLA DNA (20 μg) was linearized with Xho I and incubated with 10 units of exonuclease III (Bethesda Research Laboratories, BRL) in 250 μl of 30 m*M* tris-HCl, *p*H 8.0, 10 m*M* β-mercaptoethanol, and 2 m*M* $MgCl_2$, which had been warmed at 37°C. Portions (60 μl) were removed at 45, 120, 180, and 300 seconds into tubes containing 1 μl of 200 m*M* EDTA, *p*H 8.0, and then incubated at 65°C for 5 minutes. The samples were then diluted to 600 μl with (final concentrations) 50 m*M* sodium acetate, *p*H 4.5, 1 m*M* $ZnSO_4$, and 0.2*M* NaCl and digested with 600 units of S1 nuclease (BRL) for 30 minutes at 20°C. Digestion was stopped by adjusting the samples to 100 m*M* tris-HCl, *p*H 8.0, 10 m*M* EDTA, and 0.1 percent sodium dodecyl sulfate. The samples were extracted first with phenol and then with ether, and were then precipitated with ethanol and resuspended in 10 m*M* tris-HCl, *p*H 8.0, and 1 m*M* EDTA. One microgram of the DNA was then incubated in 25 μl of 50 m*M* tris-HCl, *p*H 8.0, 6 m*M* $MgCl_2$, 6 m*M* β-mercaptoethanol, 20 μ*M* EDTA, 50 μg of bovine serum albumin per milliliter, and all four deoxyribonucleoside triphosphates (each at 50 μ*M*), with 1.5 units of T4 DNA polymerase (BRL) for 20 minutes at 14°C and then for 5 minutes at 65°C. Five microliters of the blunt-end DNA were circularized (16 hours at 14°C) with 1 unit of T4 DNA ligase (BRL) after dilution to 20 μl and adjusting the concentrations to 20 m*M* tris-HCl, *p*H 8.0, 10 m*M* $MgCl_2$, 10 m*M* β-mercaptoethanol and 50 μ*M* adenosine triphosphate (ATP). The ligase reaction was diluted with 0.1*M* tris-HCl, *p*H 8.0, and used to transfect *E. coli* C600 (*39*). Approximately 20 colonies per nanogram of DNA were obtained. Colonies were selected at random from plates representing the different times of exonuclease III digestion and grown as small cultures. DNA from clear lysates was analyzed by digestion with restriction endonucleases Hpa II and Pvu II. From 180 colonies analyzed we selected for further study 32 colonies that had lost the Hpa II site at position −50, but retained the Pvu II site at position +33 (line b; 31 of these came from the 180-second exonuclease III digestion and one from the 300-second digestion). The DNA of some of these in vitro deletion mutants was sequenced from the Pvu II site at position +33 (line b) according to the Maxam and Gilbert technique (*40*). The sequences of eight of these mutants are aligned in part c below the wild-type adenovirus-2 DNA sequence between position −10 and −53 upstream from the mRNA start point (+1 in lines a and b) of the major late transcription unit (*12*). Numbers in parentheses at the left correspond to the position of the last base pair before the deletion which is indicated by a vertical line. The sequence replacing the wild-type sequence is underlined by a dotted line.

was the establishment in 1978 by Wu (*20*) and by Birkenmeier *et al.* (*21*) of accurate cell-free transcription systems for viral and cloned cellular genes transcribed in vivo by RNA polymerase C. In these systems the necessary factor, or factors, lacking in the purified RNA polymerase C, are supplied by a cytoplasmic fraction of KB cells (*20*) or by a nuclear extract of *Xenopus* oocytes (*21*). Unexpectedly, the groups of Brown (*18, 21*), Roeder (*21a*), and Birnstiel (*19*) found that the essential information for 5*S* RNA and tRNA transcription by RNA polymerase C is contained in an intragenic control region, in a position strikingly different from that of promoter regions in prokaryotes.

These observations raised the question whether all eukaryotic promoters are similarly located or whether this location is particular to genes transcribed by RNA polymerase C. Subsequently, Weil *et al.* (*22*) found that a system similar to that of Wu can also be used as a source of factors to promote accurate initiation of transcription by purified RNA polymerase B at the major late adenovirus-2 promoter. Briefly, these workers used a "truncated template" assay which contains, in addition to a cytoplasmic KB cell extract (S100) and RNA polymerase B, a restriction enzyme-cut DNA fragment containing, at a well-mapped position, the promoter region of the major late adenovirus-2 transcription unit. The RNA's synthesized in vitro are labeled with radioactive nucleoside triphosphates and then separated by gel electrophoresis. RNA products of discrete sizes are produced by "runoff" termination whenever specific initiation occurs. From the length of the "runoff" transcripts the position of the region coding for the 5′ end of the in vitro synthesized RNA's can be deduced and compared with the position of the 5′ end of the in vivo transcription unit. Sequence analysis of the capped in vitro synthesized 5′ terminus has verified the accuracy of initiation in vitro (*22*). Manley *et al.* (*22a*) have recently described a second system which will direct specific transcription in vitro. This system consists of a whole cell extract that does not require exogenous RNA polymerase B.

We have recently completed a study of the anatomy of the cloned chicken ovalbumin and conalbumin genes and of their transcription units (*14-16*), and have used the S100 system to demonstrate specific in vitro initiation of transcription of these cellular genes at the sites corresponding to the 5′ terminal nucleotide of the in vivo primary transcripts (*13*). Furthermore, we showed by specific fragmentation of the conalbumin gene DNA that a short segment, situated between positions −8 to −44 upstream from the initiation site and containing the "TATA" region of homology, was required for specific in vitro transcription. These results suggested to us that, in contrast to the case of RNA polymerase C, the promoter sequences for RNA polymerase B are located upstream from the mRNA start points, as in prokaryotes. (Although we have no evidence that RNA polymerase actually binds to these sequences, for the sake of convenience we use the term "promoter" in this article to designate these upstream sequences that direct specific initiation of transcription in vitro.)

Additional support for the promoter role of these sequences was provided by comparison of the transcriptional efficiencies of conalbumin and ovalbumin genes with adenovirus-2 early (E1A) and major late genes. It is noteworthy that the conalbumin and adenovirus-2 major late genes, which share an extensive 12-bp homology in their TATA box regions (*15*), are transcribed in vitro with the same efficiency. These genes were more strongly transcribed than those of adenovirus E1A and ovalbumin, which differ in the sequence of their TATA box (*16*). These observations suggest the existence of promoter sequences with different strengths (*13*). In this respect it is interesting to recall that base changes in the Pribnow box (including A → T and T → A) have marked effects on the efficiency of prokaryotic promoters.

In this article we describe in vitro genetic experiments designed to define the minimum sequence necessary to promote specific transcription in vitro by RNA polymerase B. To this end, deletion mutants of either conalbumin or

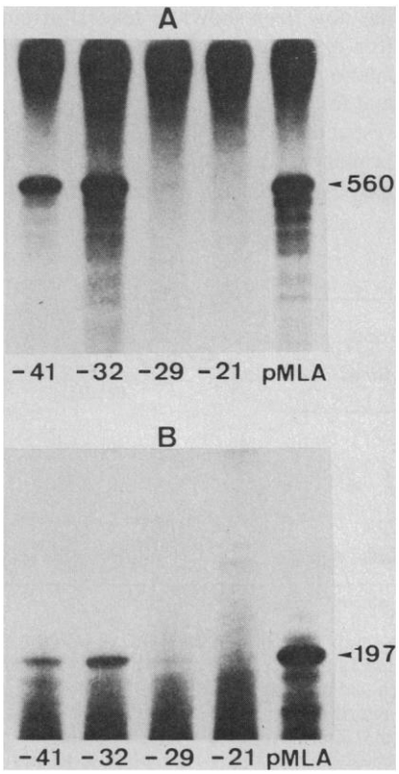


Fig. 2. (A) Electrophoretic analysis of RNA synthesized on deletion mutant DNA templates. RNA was synthesized essentially as described in Wasylyk *et al.* (*13*), for 60 minutes at 25°C in a standard reaction mixture (50 μl) containing (final concentrations) 25 μ*M* α-^{32}P-labeled cytidine triphosphate (CTP) (8000 count/min-pmole), 500 μ*M* ATP, 500 μ*M* guanosine triphosphate (GTP), 500 μ*M* uridine triphosphate (UTP), 10 m*M* tris-HCl, *p*H 7.9, 7.5 m*M* $MgCl_2$, 50 m*M* KCl, 10 percent glycerol, 0.25 m*M* dithiothreitol, 0.120 unit of calf thymus RNA polymerase B fraction PCI (*41*), 25 μl of HeLa cell S100 extract, and 1 μg of deletion mutant DNA or wild-type DNA (pMLA) linearized with restriction endonuclease Sma I (see Fig. 1). After synthesis the reaction mixture was processed and the RNA was analyzed on a 5 percent acrylamide-urea gel as described (*13*). The DNA template used for RNA synthesis is indicated by the numbers below each track. These numbers refer to the final remaining nucleotide before the deletion end point (Fig. 1, part c). RNA sizes were calculated relative to 5′ end ^{32}P-labeled Hpa II fragments of pBR322. The arrowheads point to the position of the 560-nucleotide-long RNA species. (B) Reverse transcriptase mapping of the 5′ end of RNA synthesized in vitro on deletion mutant DNA's. RNA was synthesized as above in a twofold standard reaction and, after phenol extraction, the final ethanol pellet (*13*) was resuspended in 10 μl of H_2O, and 5 μl was used to verify RNA synthesis on a 5 percent acrylamide-urea gel. The remaining 5 μl was used as the template in a reverse transcriptase reaction (*26*). The primer for the reaction was the Hha I–Hind III fragment from +80 to +197 (Fig. 1a) labeled at the Hind III end with γ-^{32}P-labeled ATP and T4 polynucleotide kinase (BRL). Approximately 0.6 pmole (100,000 count/min) of the primer was boiled in a volume of 1 μl in a tightly closed tube for 90 seconds and then immediately plunged into liquid nitrogen. Five microliters of the in vitro synthesized RNA and 2 μl of 0.25*M* tris-HCl, *p*H 7.9, 0.7*M* KCl, and 30 m*M* $MgCl_2$ were added to the frozen primer. After incubation for 10 minutes at 68°C, 1 μl of a solution containing the four deoxyribonucleoside triphosphates (each at 5 μ*M*) and 0.25*M* β-mercaptoethanol, and 1 μl of reverse transcriptase (8 units) were added. The reaction was stopped after 1 hour at 37°C by the addition of 1 μl of 1*M* NaOH, incubated at 22°C for 30 minutes, made 5*M* in urea, boiled for 1 minute, quickly chilled on ice and electrophoresed on an 8 percent acrylamide-urea gel (*40*). Numbers below each track and size markers to calibrate the gel were as described in (A). The arrowheads point to the extended DNA primer which is 197 nucleotides long. Sizes were calculated from comparison with Hpa II pBR322 fragments run on the same gel.

adenovirus major late genes [because of their 5′ end upstream sequence homologies (*15*), we have used these two genes interchangeably] have been constructed in vitro, propagated as plasmid DNA, and used as templates for in vitro transcription reactions. Our studies demonstrate that a region between positions −12 and −32 upstream from the mRNA start points is essential to promote specific transcription in vitro.

Mapping of the Promoter 5′ Boundary

To determine the promoter 5′ boundary we constructed a series of deletion mutants approaching the 5′ end of the adenovirus major late transcription unit. Plasmid pMLA (Fig. 1a), which contains the major late promoter, was cleaved at the unique Xho I site at position −250 and the ends of the linear DNA were trimmed back with a combination of exonuclease III and S1 nuclease (*18*). The resulting family of shortened linear DNA molecules was circularized with DNA ligase and used to transform *Escherichia coli*. Plasmids containing deletions centered on the Xho I site were first screened by restriction enyzme analysis. Because we knew (*23*) that pMLA digested by Hpa II is an efficient template for transcription in vitro we chose to analyze only those deletions which extended in the 3′ direction beyond the Hpa II site at −50, but not as far as the Pvu II site at +33 (Fig. 1b). The sequences of eight deletion mutants selected for further study are shown in Fig. 1c. The dotted lines underline sequences which originate upstream from the Xho I site and which replace the sequences deleted between the Xho I site and the mRNA start point. These sequences come from the 3′ end of the adenovirus IVa_2 gene which is immediately 5′ to the major late promoter and has the opposite polarity (*24*).

To determine whether the deletions containing recombinants are specifically transcribed, we used each mutant as a template for transcription in vitro (Fig. 2A). Using Sma I linearized DNA as template (Fig. 1A), S100 extract from uninfected HeLa cells, and purified calf thymus polymerase B we obtained the RNA transcripts shown in Fig. 2A. The band at 560 nucleotides represents the runoff product of specific transcription in vitro. Bands below 560 are due to premature termination (*22, 25*) while the smear at the top of the gel is due to nonspecific initiation by both RNA polymerase B and the endogenous RNA polymerase C present in the S100 extract. Plasmids with deletions ending 5′ to position −32 are specifically transcribed with efficiencies similar to that of pMLA (only those corresponding to positions −41 and −32 are shown in Fig. 2). The plasmid with a deletion to position −29 gives a very faint band of the correct size, but at most 1/100 the intensity of the parental pMLA. We have not been

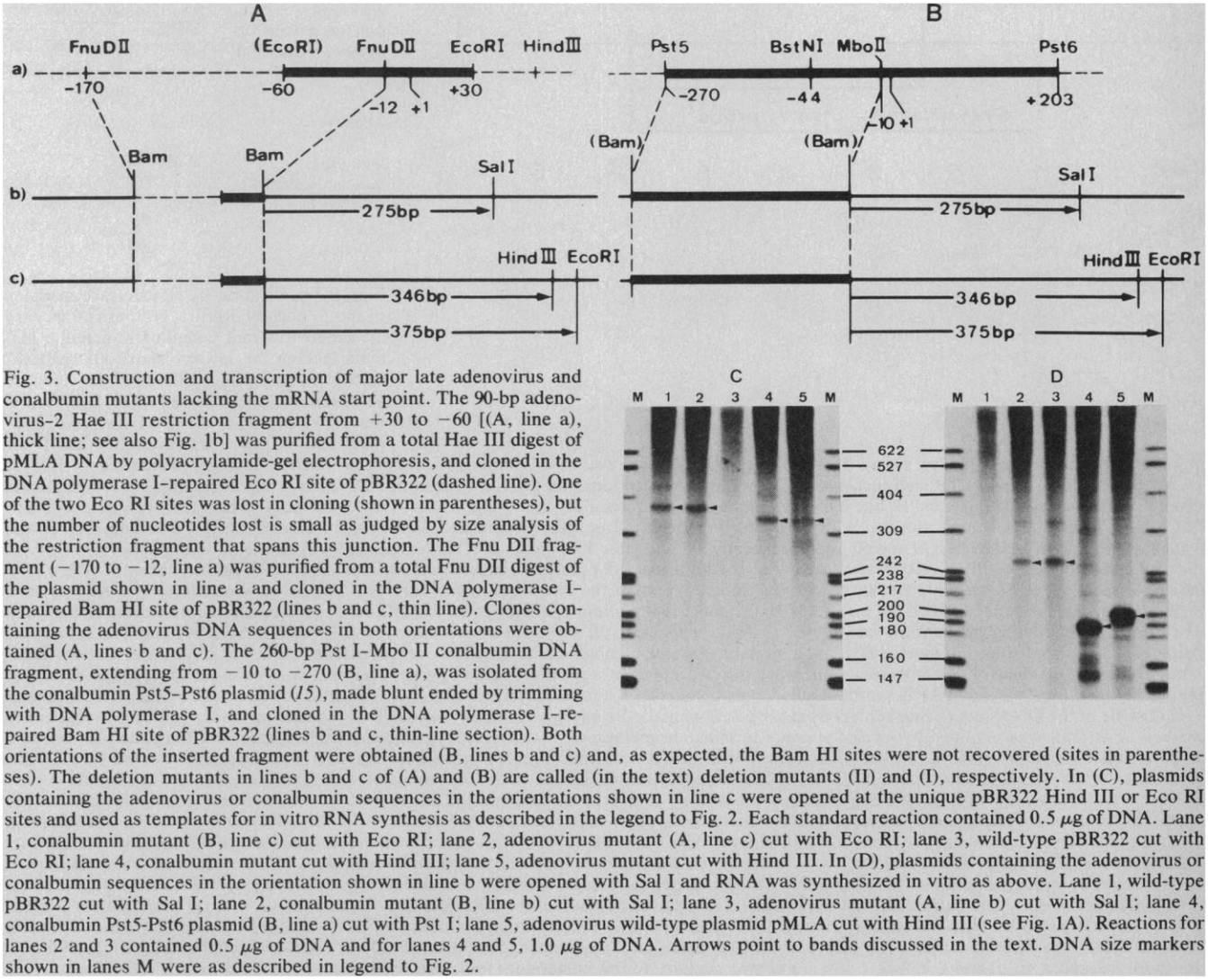


Fig. 3. Construction and transcription of major late adenovirus and conalbumin mutants lacking the mRNA start point. The 90-bp adenovirus-2 Hae III restriction fragment from +30 to −60 [(A, line a), thick line; see also Fig. 1b] was purified from a total Hae III digest of pMLA DNA by polyacrylamide-gel electrophoresis, and cloned in the DNA polymerase I-repaired Eco RI site of pBR322 (dashed line). One of the two Eco RI sites was lost in cloning (shown in parentheses), but the number of nucleotides lost is small as judged by size analysis of the restriction fragment that spans this junction. The Fnu DII fragment (−170 to −12, line a) was purified from a total Fnu DII digest of the plasmid shown in line a and cloned in the DNA polymerase I-repaired Bam HI site of pBR322 (lines b and c, thin line). Clones containing the adenovirus DNA sequences in both orientations were obtained (A, lines b and c). The 260-bp Pst I-Mbo II conalbumin DNA fragment, extending from −10 to −270 (B, line a), was isolated from the conalbumin Pst5-Pst6 plasmid (*15*), made blunt ended by trimming with DNA polymerase I, and cloned in the DNA polymerase I-repaired Bam HI site of pBR322 (lines b and c, thin-line section). Both orientations of the inserted fragment were obtained (B, lines b and c) and, as expected, the Bam HI sites were not recovered (sites in parentheses). The deletion mutants in lines b and c of (A) and (B) are called (in the text) deletion mutants (II) and (I), respectively. In (C), plasmids containing the adenovirus or conalbumin sequences in the orientations shown in line c were opened at the unique pBR322 Hind III or Eco RI sites and used as templates for in vitro RNA synthesis as described in the legend to Fig. 2. Each standard reaction contained 0.5 μg of DNA. Lane 1, conalbumin mutant (B, line c) cut with Eco RI; lane 2, adenovirus mutant (A, line c) cut with Eco RI; lane 3, wild-type pBR322 cut with Eco RI; lane 4, conalbumin mutant cut with Hind III; lane 5, adenovirus mutant cut with Hind III. In (D), plasmids containing the adenovirus or conalbumin sequences in the orientation shown in line b were opened with Sal I and RNA was synthesized in vitro as above. Lane 1, wild-type pBR322 cut with Sal I; lane 2, conalbumin mutant (B, line b) cut with Sal I; lane 3, adenovirus mutant (A, line b) cut with Sal I; lane 4, conalbumin Pst5-Pst6 plasmid (B, line a) cut with Pst I; lane 5, adenovirus wild-type plasmid pMLA cut with Hind III (see Fig. 1A). Reactions for lanes 2 and 3 contained 0.5 μg of DNA and for lanes 4 and 5, 1.0 μg of DNA. Arrows point to bands discussed in the text. DNA size markers shown in lanes M were as described in legend to Fig. 2.

able to detect a runoff transcript of the correct size from the plasmid with a deletion to position −21.

To verify that the transcripts shown in Fig. 2A actually start at the previously mapped adenovirus major late start point, we used reverse transcriptase primer extension mapping (*26*) to locate the 5′ end of RNA synthesized in vitro. A primer consisting of the Hha I–Hind III fragment (+80 to +197, Fig. 1b) labeled with ^{32}P at the Hind III site was annealed to the RNA synthesized in vitro and extended with reverse transcriptase. The size of the 197-base DNA band (Fig. 2B) is in excellent agreement with the size expected for a reverse transcript synthesized on an RNA template with a 5′ end corresponding to the sequenced start point (Fig. 1a) (*12*). The relative intensity of the 197-base band also agrees with the intensity of the runoff transcript for each template. Together, these results show that DNA 5′ to position −32 can be replaced without affecting specific transcription in vitro. Removal of the three nucleotides −32, −31, and −30 dramatically reduces transcription, thus locating the 5′ boundary of an indispensable part of the major late adenovirus promoter to the base pairs −32, −31, or −30 from the mRNA start point.

Mapping of the Promoter 3′ Boundary

To determine whether sequences located around position −30 can promote specific transcription in vitro even when separated from the natural mRNA start point, we constructed recombinant plasmids containing pBR322 sequences in place of the adenovirus and conalbumin mRNA start points. An adenovirus recombinant lacking the mRNA start point was prepared by cloning a 90-bp adenovirus Hae III restriction fragment (from position −60 to +30, Fig. 1b) into pBR322 (Fig. 3A, line a). From this recombinant a 158-bp Fnu DII fragment, extending from nucleotide −12 in adenovirus DNA (Fig. 3A, line a), was isolated and cloned in the Bam HI site of pBR322. Recombinants containing the adenovirus DNA sequences in both orientations were isolated (Fig. 3A, lines b and c). Similar recombinants were constructed (Fig. 3B, lines c and d) for the conalbumin gene by using a 260-bp conalbumin DNA fragment, extending from the Pst I site (Pst5) at −270 to the Mbo II site at −10 (Fig. 3B, line a). Recombinants corresponding to the orientation shown in lines b and c were called deletion mutants (II) and (I) (Fig. 4), respectively. These adenovirus and conalbumin recombinants contain the pBR322 sequences adjacent to the Bam HI site in the place of the sequences downstream from position −12 of the adenovirus gene, and −10 of the conalbumin gene. The junctions of these

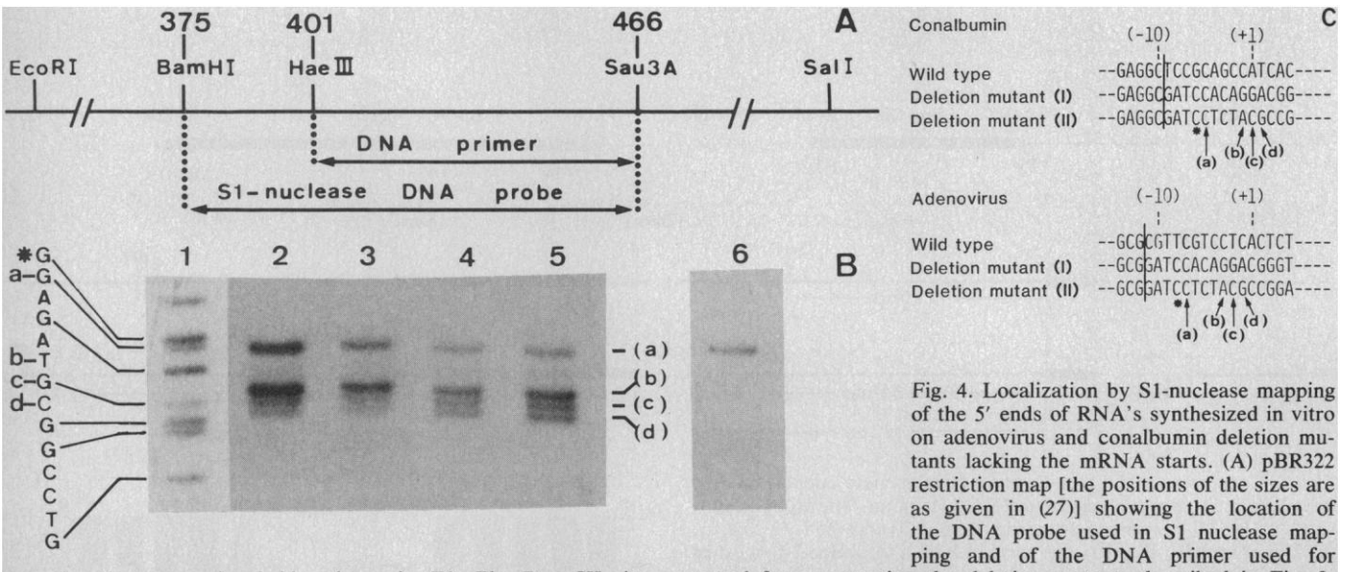


Fig. 4. Localization by S1-nuclease mapping of the 5′ ends of RNA's synthesized in vitro on adenovirus and conalbumin deletion mutants lacking the mRNA starts. (A) pBR322 restriction map [the positions of the sizes are as given in (*27*)] showing the location of the DNA probe used in S1 nuclease mapping and of the DNA primer used for preparing the sequencing ladder shown in (B). The Bam HI site was used for constructing the deletion mutants described in Fig. 3, A and B. (B) RNA was synthesized in vitro (twofold standard reactions) on the Sal I linearized DNA from the adenovirus and conalbumin deletion mutants (see Fig. 3, A and B, line b) processed as described in the legend to Fig. 2, dissolved in 100 μl of 50 m*M* tris-HCl, *p*H 7.9, 2 m*M* $CaCl_2$, and 10 m*M* $MgCl_2$, and digested with 50 μg/ml ribonuclease-free deoxyribonuclease (*42*) for 10 minutes at 37°C. Sodium-EDTA and proteinase K were added to 15 m*M* and 50 μg/ml, respectively, and after 30 minutes at 37°C the reaction was extracted with phenol (*p*H 8.0) and then by chloroform. The RNA was ethanol precipitated with about 0.5 pmole of single-stranded Bam HI–Sau 3A probe [see part (A)], 5′ end labeled with ^{32}P (specific activity, 500,000 count/min-pmole, 5′ end) at the Sau 3A site (position 466). The pellet was redissolved in 35 μl of 50 percent formamide, 40 m*M* PIPES, *p*H 6.5, 0.4*M* NaCl, and 1 m*M* sodium-EDTA, heated to 85°C for 5 minutes and incubated for 12 hours, at 42°C. After rapid cooling the mixture was diluted tenfold, adjusted to 30 m*M* sodium acetate (*p*H 4.3), 3 m*M* $ZnSO_4$, 0.4 m*M* NaCl, and, per milliliter, 7 μg of sonicated calf thymus DNA; the mixture was then incubated with 4000 units of S1 nuclease (Miles) at 25°C. After 4 hours (lanes 3 and 4) and 8 hours (lanes 2 and 5) the reactions were stopped and electrophoresed on an 8 percent acrylamide-urea sequencing gel, as described previously (*13*). Lanes 2 and 3: RNA synthesized on conalbumin DNA template. Lanes 4 and 5: RNA synthesized on adenovirus DNA template. Lane 1. One of the DNA sequencing ladders used for positioning the S1 nuclease-resistant bands. This sequencing ladder (which corresponds to the coding strand) was obtained by using the chain termination method (*43*) with dideoxyguanosine triphosphate; the template was the conalbumin deletion mutant (II) (Fig. 3B, line b), linearized with Eco RI and limit-digested with exonuclease III, and the primer was the Hae III–Sau 3A fragment [see (A)] 5′ end labeled with ^{32}P at the Sau 3A site. Lane 6. Same as lane 5 but α-amanitin (1 μg/ml) was present during the in vitro RNA synthesis. (C) Sequences (noncoding strand) of the wild type, and the conalbumin and adenovirus deletion mutants lacking the mRNA start point. Deletion mutants (I) and (II) correspond to the mutants shown in lines c and b, respectively, of Fig. 3A (adenovirus) and 3B (conalbumin). The wild-type conalbumin sequence (noncoding strand) is from Cochet *et al.* (*15*). The wild-type adenovirus major late sequence was taken from Ziff and Evans (*12*) and has been confirmed for the adenovirus-2 used in our laboratory. The DNA sequences of the conalbumin mutants I and II were verified by sequencing. The DNA sequences of the adenovirus mutants I and II were deduced from the regeneration of the Bam HI sites (Fig. 3A, lines b and c). The nucleotides are numbered relative to the position of the mRNA start point (+1) in the wild-type and mutant DNA's. The G marked by an asterisk in the coding strand sequence ladder shown in (B) is complementary to the C's marked by asterisks in the noncoding DNA strands of the deletion mutants (II) shown in (C) (see text). The 3′ terminal nucleotide of the DNA bands marked a to d in (B), lanes 2 to 5, are those marked a, b, c, and d in lane 1 and are complementary to the nucleotides indicated by the arrows (a to d) in (C).

plasmids (vertical lines in Fig. 4C) were verified by sequence analysis for the conalbumin recombinants and deduced from regeneration of the Bam HI sites for the adenovirus recombinants (see legend to Fig. 4).

The conalbumin and adenovirus recombinants lacking the natural mRNA start points were transcribed in vitro. Adenovirus and conalbumin deletion mutants (I) (Fig. 3, A and B, line c) were linearized with either Hind III or Eco RI, and after transcription in vitro the runoff RNA products were analyzed by gel electrophoresis. RNA's of about 365 (Fig. 3C, lanes 1 and 2) and 335 (Fig. 3C, lanes 4 and 5) nucleotides were observed for the Hind III and Eco RI linearized recombinants, corresponding to the sizes expected if transcription initiates in pBR322 DNA close to the "Bam HI junction" and terminates at the end of the DNA fragments. No RNA of equivalent size was observed when "wild-type" pBR322 DNA linearized with Eco RI was used as template (Fig. 3C, lane 3). Deletion mutants (II) containing adenovirus or conalbumin sequences in the opposite orientation (Fig. 3, A and B, line b) and linearized with Sal I gave runoff transcripts of about 265 nucleotides (Fig. 3D, lanes 2 and 3), again as expected if transcription starts in pBR322 sequences and terminates at the end of the DNA fragment. No RNA's of similar size were observed with a Sal I linearized wild-type pBR322 template (Fig. 3D, lane 1). To compare the template efficiencies of the mRNA start point deletion mutants with the intact parental sequences, the optimal DNA concentrations for specific transcription were determined for each different template by varying their concentration in the incubation medium from 10 to 100 micrograms per milliliter. Under optimal conditions for wild-type adenovirus pMLA DNA linearized with Hind III (Fig. 1, line b) and conalbumin Pst5-Pst6 plasmid linearized with Pst I (Fig. 3B, line a) specific runoff transcripts were observed (arrowheads in Fig. 3D, lanes 4 and 5). At optimal DNA concentrations the intensity of the specific transcript bands for the deletion mutants is about 10 percent of the wild type (arrowheads in Fig. 3D, lanes 2 and 3).

These results indicate that the sequences located upstream from positions −10 and −12 in the conalbumin and major adenovirus late transcription units can direct specific initiation of transcription when fused to pBR322 sequences. To determine where in the pBR322 sequences these transcripts originate, the 5′ ends of the RNA's made in vitro were

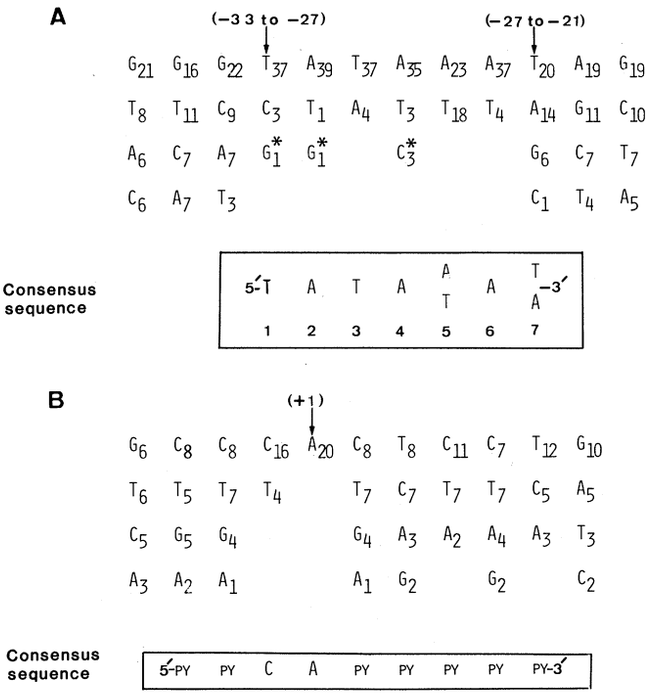


Fig. 5. Consensus sequences around the 5′ end of protein-coding genes. (A) The DNA sequences (noncoding strand) of 41 genes [for references, see (*11*)] have been compared upstream from the mRNA start points and aligned to yield maximum homology in the −32 to −21 region. The numbers in parentheses and the arrows indicate the position of the first and last bases of the TATA box with respect to the mRNA start points. The subscripts to the bases correspond to the frequency of a particular residue at each position. Bases marked with an asterisk are present in some histone genes (*44*). (B) The DNA sequence (noncoding strand) of 20 genes for which the mRNA start point has been established [for references, see (*11*)] were aligned with respect to the base corresponding to the 5′ end of the mRNA (+1 and arrow). Subscripted numbers are as in (A) (*PY*, pyrimidine).

located by S1 nuclease mapping. A single-stranded DNA probe labeled with ^{32}P at the pBR322 Sau 3A site (Fig. 4A) [see (*27*) for map coordinates] was prepared (see legend to Fig. 4B). RNA was synthesized in vitro on the DNA of adenovirus and conalbumin mRNA start point deletion mutants (II) which had been linearized with Sal I (Fig. 3, A and B, line b). This RNA was hybridized to the labeled DNA probe and, after S1 nuclease digestion, the S1 resistant DNA was analyzed by electrophoresis alongside DNA sequence ladders (see legend to Fig. 4, A and B).

Four major S1 resistant products (labeled a to d in Fig. 4B) were obtained with RNA synthesized on the deletion mutant (II) templates. We observed little change in the relative intensities of the four bands after either 4 or 8 hours of S1 nuclease digestion for the conalbumin (Fig. 4B, lanes 3 and 2) or adenovirus DNA templates (Fig. 4B, lanes 4 and 5). Furthermore, the same four bands were observed when the resistant RNA-DNA hybrids were purified and redigested for a further 4 hours with freshly added S1 nuclease (not shown).

From the position of the S1 nuclease resistant bands (b, c, and d) present in lanes 2 to 5, one can infer the position of the 5′ end of the corresponding RNA species (assuming that S1 nuclease makes blunt-end cuts at the ends of the DNA : RNA hybrid molecules). Bands b, c, and d, migrate, as shown in the sequencing ladder (lane 1), at the positions corresponding to the bases T, G, and C. Since the sequencing ladder corresponds to the coding strand, the 5′ terminal nucleotides of the RNA species responsible for bands b, c, and d, are A, C, and G, respectively. This is illustrated in Fig. 4C, where arrows b, c, and d indicate, in the noncoding DNA strand, the bases which correspond to the 5′ terminal nucleotides of the RNA species synthesized in vitro on the conalbumin and adenovirus deletion mutants (II). Band a present in lanes 2 to 5 migrates at the position corresponding to the undigested DNA probe and is also observed when RNA is synthesized in the presence of α-amanitin (1 μg/ml), which inhibits RNA polymerase B (Fig. 4B, lane 6). Hence, this band may result from RNA transcripts made with RNA polymerase C or, less likely, could be due to an incomplete separation of the probe DNA from its complementary DNA strand during its preparation. In any case, since band a is not the major product when either RNA synthesized on conalbumin (Fig. 4B, lanes 2 and 3) or adenovirus (Fig. 4B, lanes 4 and 5) mutant DNA was used for S1 nuclease mapping, and since there are no bands visible on the autoradiogram below band d, we conclude that most of the RNA's synthesized in vitro by RNA polymerase B have their 5′ ends as indicated by arrows b, c, and d in Fig. 4C. The lengths of the in vitro runoff transcripts observed in

Fig. 3 are entirely consistent with start points at these positions.

From the S1 nuclease mapping results it appears that the sequences located upstream from positions −10 and −12 for conalbumin and adenovirus, respectively, can direct RNA polymerase B to initiate in pBR322 sequences close to the positions corresponding to the wild-type start points (position +1 in Fig. 4C). From the relative intensity of bands b, c, and d in Fig. 4B, it appears that RNA synthesis may start preferentially in vitro at the position corresponding to the A located at −1 and −3 in the conalbumin and adenovirus deletion mutant (II), respectively. This preference might reflect some sequence requirement at the start point. In this respect, it is interesting that both conalbumin and adenovirus major late in vitro transcripts are initiated with A and, more generally, mRNA's are preferentially initiated with an A which is surrounded in the noncoding DNA strand by pyrimidines (Fig. 5B).

In vitro, however, initiating with an A may not be an absolute requirement, since the presence of the fainter S1 nuclease bands (c) and (d) suggests that initiation of transcription may also be taking place at the positions corresponding to the neighboring C and G (arrows c and d in Fig. 4C). This is particularly noticeable in the case of adenovirus where the initiating A in vitro is three nucleotides upstream from the position corresponding to the wild-type start point (position +1 in Fig. 4C). However, it should be stressed that the present identification of the in vitro start points on the deletion mutants (II) is based on the assumption that S1 nuclease makes blunt-end cuts at the ends of the DNA : RNA hybrids. Although our results demonstrate clearly that initiation of RNA synthesis is taking place on the deletion mutants (II) at sites very close to the wild-type start points, direct sequencing of the 5′ ends of the in vitro synthesized RNA's is required to establish definitely multiplicity and exact locations of the in vitro start points.

Role of TATA Box and Start Point in Specific Initiation in vitro

Taken together, the results of the experiments described in this article suggest that the sequences from −10 to −44 for the conalbumin gene and −12 to −32 for the adenovirus major late transcription unit play an essential role in the specific transcription of these genes in vitro by RNA polymerase B. Within this region adenovirus and conalbumin share the sequence: 5′-CTATAAAAGGGG-3′, where the 5′-C is at position −32 and the 3′-G is at position −21 in both genes (*15*). This region contains the TATA box homology (*14*) which has been found upstream from the mRNA start points of all sequenced eukaryotic genes transcribed by RNA polymerase B with the exception of the papovavirus late genes and the adenovirus early region 2 (*28*). The consensus TATA box sequence which is shown in Fig. 5A is compiled from sequences at the 5′ end of 41 eukaryotic genes and shows a striking similarity to the bacterial Pribnow box. Among the individual eukaryotic genes, adherence to the consensus sequence is not absolute. While positions 1 to 4 and 6 have a single predominant nucleotide, positions 5 and 7 are represented equally by A and T (see Fig. 5A). No position in the TATA box is invariant, although in 39 of 41 cases A is found in position 2. The exact location of the TATA box varies from gene to gene with the T in position 1 falling between positions −33 and −27 from the mRNA start point (however, for 30 of the 41 genes the T in position 1 falls within one base of position −31). Comparison of the sequences flanking the TATA box within 10 bp does not reveal any obvious additional homologies, although these regions tend to be purine-rich on the noncoding strand. Furthermore, it is not possible to form secondary structures common to all genes in the sequences surrounding the TATA box.

Several lines of evidence suggest that the TATA box is the element present in the −32 to −12 region which is indispensable for the initiation of specific transcription in vitro. First, deletion to position −29 of the adenovirus upstream sequence (Fig. 1c) which removes the TA in positions 1 and 2 of the TATA box (Fig. 5A, *consensus sequence*) results in at least a 100-fold decrease in the efficiency of specific transcription. In contrast, when the deletion end point is three bases upstream, at position −32, no significant change in the efficiency of specific in vitro transcription is observed. This result, however, must be interpreted with some caution. Indeed, in deletion mutants, the DNA upstream from the deletion end points is different from the original DNA. Therefore, we cannot rule out the possibility that the effects we observe are due to the sequences drawn close to the promoter through the process of deletion. More recently we have obtained unequivocal evidence that the TATA box is implicated in the promotion of transcription. A single base change which converts to a G the T in position 3 of the conalbumin TATA box has been obtained through in vitro site-directed mutagenesis. When this conalbumin mutant gene is transcribed in vitro the efficiency is at most 10 percent of wild type (*29*). This down mutation points to the TATA box as being essential for specific initiation in vitro. Furthermore, we have shown that a cloned −12 to −32 fragment of the adenovirus major late gene is sufficient to promote specific initiation of transcription in vitro (*29a*).

Is, in fact, the TATA box part of a promoter region—that is, a region of the DNA that participates in initiation of transcription and to which RNA polymerase binds? From the evidence we have obtained in vitro this region seems to be indispensable for the initiation of transcription and thus fits the original genetic definition of a promoter. However, our results do not indicate whether or not RNA polymerase B binds to the TATA box region. Studies of the interaction of RNA polymerase and factors present in the S100 extract with the DNA in this region will be necessary to answer this question.

The situation for initiation of transcription by RNA polymerase B is in marked contrast with what has been observed for genes transcribed by RNA polymerase C, where an essential sequence component directing initiation of transcription is located within the genes—that is, downstream from the start points. However, the reduced template efficiency of the deletion mutants which lack the mRNA start points suggests the involvement of sequences other than the TATA box in specific in vitro initiation of transcription by RNA polymerase B. That the role of the sequences downstream from position −10 in conalbumin and −21 in adenovirus is different from the role of the TATA box is clearly seen from the inability of these regions themselves to direct specific initiation [see (*13*) and Fig. 2]. From the recent work of Hu and Manley (*30*) we conclude that the downstream sequences important for efficient transcription in vitro seem to be located between positions −12 and +3, since these authors found that an adenovirus major late gene mutant, lacking the sequence downstream from position +3, is transcribed in vitro as efficiently as the wild-type DNA. The sequences surrounding the mRNA start points of 20 RNA polymerase B genes for which the 5′ end of the mRNA is accurately known were used to derive the consensus sequence shown in Fig. 5B. The striking characteristic of this sequence is that only the A at the mRNA start point and the preceding C

seem to be conserved (this consensus is certainly biased, however, since up to now only A-containing mRNA 5′ ends have been accurately mapped). Rather than having a particular sequence the mRNA start point consists of an A residue surrounded by pyrimidines and situated at 27 to 33 bp from the TATA box, about three turns of the DNA helix.

It is striking that the preferential start points used in the deletion mutants analyzed in this article appear also to be an A residue surrounded by pyrimidines. From our present results we cannot decide whether the decrease in efficiency of transcription of these deletion mutants is due to the change in distance between the TATA box and the new start points or to the particular sequence surrounding these start points, or to both. It is likely that a set of rules comprised of both these features as well as sequence considerations in the TATA region determines the overall promoter efficiency. That we see transcription at all on our start-point deletion mutants indicates that a degree of flexibility exists in the mechanism of start-point selection.

Is There More to the RNA Polymerase B Promoter?

In the in vitro specific transcription system, we have clearly established that the TATA box and the mRNA start point play an important role, possibly as a promoter sequence, in directing the specificity and efficiency of transcription. Whether these sequences play a similar role in vivo has not been definitively answered. It is possible that other sequences which may lie upstream from the TATA box or downstream from the mRNA start point play an important role in vivo. That the present in vitro system is very inefficient in promoting specific transcription (*13*) could be due to a lack of important factors in the extracts which may interact with sequences other than the TATA and mRNA start-point sequences.

There are several reasons for believing that other sequences could be part of the in vivo promoter. In the upstream sequences of RNA polymerase B genes, there are homologies apart from the TATA box and the mRNA start point. A model eukaryotic sequence 5′-GC$^{C}_{T}$CAATCT-3′ can be drawn for the −70 to −80 positions for a variety of eukaryotic gene sequences, and other homologies have been noted (*15, 16, 31*). Further evidence comes from two experiments in which deletion mutants in upstream sequences were prepared and the expression of the DNA was studied in vivo. First, Grosschedl and Birnstiel (*32*) found that, for the sea urchin histone *H2A* gene injected into *Xenopus* oocytes, deletion of the TATA sequence did not abolish transcription. Instead, a number of new start points of lower efficiency were generated. Deleting the mRNA start point, but not the TATA box, created a new mRNA 5′ terminus (*32*). Second, Benoist and Chambon (*33*) showed that a recombinant plasmid, constructed by inserting SV40 early genes into pBR322, expressed the early genes when introduced into eukaryotic cells. A recombinant with a 60-bp deletion, which removes the mRNA start point and the TATA box, still expressed the SV40 genes, while more distal deletion of all upstream sequences led to inactivation of the early genes. Along the same lines it should be recalled that there are two notable exceptions to the universal presence of a TATA box upstream from the mRNA start point: the papovavirus late genes and the adenovirus-2 *E2* gene (*28*). These exceptions, which are accompanied by the occurrence of multiple start points in the transcription units, suggest that there might be more than one class of promoters for RNA polymerase B.

Although the experiments described here suggest that some upstream sequences could also be important to the initiation of transcription, they do not exclude the possibility that the TATA box plays an essential role in vivo as a promoter. In the experiments of Benoist and Chambon (*33*) it is possible that, in the mutant lacking the TATA box, minor promoters (initiating upstream from the major start points) were responsible for the transcription of the early genes. In the experiments of Grosschedl and Birnstiel (*32*), a deletion mutant lacking the *H2A* gene-specific conserved DNA sequence but retaining the TATA box and the mRNA start point sequences functions better than the wild-type gene. It is also worth noting that in the deletion mutant of Grosschedl and Birnstiel (*32*), which lacks the natural mRNA start point, the 5′ end of the novel mRNA maps about 24 nucleotides downstream from the TATA box, in good agreement with our experiments in vitro with the adenovirus and conalbumin deletion mutants lacking the natural sequences downstream from −12 and −10. Similar results have been obtained in vivo by Benoist and Chambon (data in preparation) with SV40 early gene deletion mutants lacking the mRNA start points, but not the TATA box. Finally, two points should be emphasized. First, the use of deletion mutants, although useful for detecting the functional sequences, could lead to artifactual results because the deleted sequences are replaced by other sequences that are not necessarily neutral with respect to initiation of transcription. Second, in vivo the transcribed DNA is probably organized into some form of chromatin structure [for a review, see (*34*)] not present in the in vitro system but may be important in the functioning of some control regions.

Several lines of evidence indicate that other important components involved in the in vivo regulation of transcription are missing in the present cell-free system and even in a semi-in vivo system such as *Xenopus* oocytes. In the in vitro cell-free system both the ovalbumin and the related *X* gene are equally poorly transcribed compared to the conalbumin gene (*13*). In contrast, the in vivo rate of ovalbumin gene transcription, under hormonal stimulation, is higher than that of the conalbumin gene (*1*) and at least 30 times that of the *X* gene (*35*). When the complete ovalbumin gene (*14*) including 0.7 kb of 5′ flanking sequences is injected into *Xenopus* oocytes, no specific transcription can be detected (*36*). It has also been found that, of the injected histone genes, only *H2A, H2B*, and *H3*, but not *H4*, are efficiently transcribed (*32, 37*), although these genes in vivo are probably expressed coordinately. It is clear that not all of the necessary components required for specific initiation of transcription are present in all cells, since Breathnach *et al.* (*38*) found that initiation of transcription takes place from an aberrant start point when the ovalbumin gene is introduced into mouse L cells.

The availability of the cell-free system in vitro, which can be dissected into its essential components and used to identify by complementation further important factors, and the possibility of obtaining large quantities of isolated genes that can be mutated in vitro, represent a large step toward the identification of the components involved in initiation of transcription. It is clear, however, that the in vitro approach must be complemented by studies in which genes altered by site-directed mutagenesis are introduced into cells (*38a*).

References and Notes

1. G. S. McKnight and R. D. Palmiter, *J. Biol. Chem.* **254**, 9050 (1979); F. Perrin, M. Cochet, P. Gerlinger, B. Cami, J. P. LePennec, P. Chambon, *Nucleic Acids Res.* **6**, 2731 (1979).
2. M. Rosenberg and D. Court, *Annu. Rev. Genet.* **13**, 319 (1979).
3. F. Jacob, A. Ullman, J. Monod, *C.R. Acad. Sci.* **258**, 3125 (1964).
4. R. Losick and M. Chamberlin, in *RNA Polymerase* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1976), pp. 285-329.

5. D. Pribnow, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 784 (1975); *J. Mol. Biol.* **9**, 419 (1975).
6. H. Schaller, C. Gray, K. Herrmann, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 737 (1975).
7. By convention we give in this article only the anti-sense (noncoding) DNA (5′ → 3′) strand and therefore transcription is proceeding from left to right. DNA sequences in the direction of transcription (downstream) are numbered with positive integers, whereas sequences 5′ to the start point (upstream) are given negative values.
8. P. Chambon, *Annu. Rev. Biochem.* **44**, 613 (1975).
9. R. G. Roeder, in *RNA Polymerase* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1976), pp. 285-329.
10. P. Chambon, *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1209 (1978).
11. R. Breathnach and P. Chambon, *Annu. Rev. Biochem.*, in press.
12. E. B. Ziff and R. M. Evans, *Cell* **15**, 1463 (1978).
13. B. Wasylyk, C. Kédinger, J. Corden, O. Brison, P. Chambon, *Nature (London)* **285**, 367 (1980).
14. F. Gannon, *et al.*, *ibid.* **278**, 428 (1979).
15. M. Cochet, F. Gannon, R. Hen, L. Maroteaux, F. Perrin, P. Chambon, *ibid.* **282**, 567 (1979).
16. C. Benoist, K. O'Hare, R. Breathnach, P. Chambon, *Nucleic Acids Res.* **8**, 127 (1980).
17. B. Sollner-Webb and R. H. Reeder, *Cell* **18**, 485 (1979).
18. S. Sakonju, D. F. Bogenhagen, D. D. Brown, *ibid.* **19**, 13 (1980); D. F. Bogenhagen, S. Sakonju, D. D. Brown, *ibid.*, p. 27; H. R. B. Pelham and D. D. Brown, *Proc Natl. Acad. Sci. U.S.A.* **77**, 4170 (1980).
19. A. Kressmann, H. Hofstetter, E. Di Capua, R. Grosschedl, M. Birnstiel, *Nucleic Acids Res.* **7**, 1749 (1979).
20. G. J. Wu, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2175 (1978).
21. E. H. Birkenmeier, D. Brown, E. Jordan, *Cell* **15**, 1077 (1978).
21a. D. R. Engelke, N. G. Sun-yu, D. S. Shastry, R. G. Roeder, *ibid.* **19**, 717 (1980).
22. P. A. Weil, D. S. Luse, J. Segall, R. G. Roeder, *ibid.* **18**, 469 (1979).
22a. J. L. Manley, A. Fire, A. Cono, P. A. Sharp, M. L. Gester, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3855 (1980).
23. J. Corden and C. Kédinger, unpublished data.
24. G. Akusjärvi and U. Pettersson, *J. Mol. Biol.* **134**, 143 (1979).
25. J. E. Darnell, Jr., in *From Gene to Protein* (Academic Press, New York, 1979), pp. 207-227.
26. P. K. Ghosh, V. B. Reddy, M. Piatak, P. Lebowitz, S. M. Weissman, *Methods Enzymol.* **65**, 580 (1980).
27. J. G. Sutcliffe, *Nucleic Acids Res.* **5**, 2721 (1978).
28. C. C. Baker, J. Heriss, G. Courtois, F. Galibert, E. Ziff, *Cell* **18**, 569 (1979).
29. B. Wasylyk, R. Derbyshire, A. Guy, D. Molko, A. Roget, R. Teoule, P. Chambon, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
29a. P. Sassone-Corsi and J. Corden, unpublished data.
30. S. L. Hu and J. Manley, personal communication.
31. T. Maniatis, E. F. Fritsch, J. Lauer, R. M. Lawn, *Annu. Rev. Genet.*, in press.
32. R. Grosschedl and M. L. Birnstiel, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1432 (1980).
33. C. Benoist and P. Chambon, *ibid.*, in press.
34. D. Mathis, P. Oudet, P. Chambon, *Prog. Nucleic Acid Res. Mol. Biol.*, in press.
35. M. LeMeur, N. Glanville, J. L. Mandel, P. Gerlinger, R. Palmiter, P. Chambon, in preparation.
36. D. Mathis and P. Chambon, in preparation.
37. M. Birnstiel, personal communication.
38. R. Breathnach, N. Mantei, P. Chambon, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 740 (1980).
38a. R. C. Mulligan and P. Berg, *Science* **209**, 1422 (1980); A. Pellicer *et al.*, *ibid.*, p. 1414.
39. P. Humphries, M. Cochet, A. Krust, P. Gerlinger, P. Kourilsky, P. Chambon, *Nucleic Acids Res.* **4**, 2389 (1977).
40. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 498 (1980).
41. C. Kédinger and P. Chambon, *Eur. J. Biochem.* **28**, 283 (1972).
42. O. Brison and P. Chambon, *Anal. Biochem.* **75**, 402 (1976).
43. A. J. H. Smith, *Methods Enzymol.* **65**, 560 (1980).
44. M. Busslinger, R. Portmann, J. C. Irminger, M. L. Birnstiel, *Nucleic Acids Res.* **8**, 957 (1980).
45. We thank the Viral Cancer Program, National Cancer Institute (Dr. Beard) for the avian myeloblastosis virus reverse transcriptase; C. Hauss, J. M. Bornert, C. Wasylyk, and K. Dott for technical assistance; and C. Kutschis and B. Chambon for help in preparing the manuscript. This investigation was supported by grants from the CNRS (ATPs 3907, 3558, 4160), INSERM (ATP 72.79.104), and the Fondation pour la Recherche Médicale Française.

7 July 1980

Altering Genotype and Phenotype by DNA-Mediated Gene Transfer

Angel Pellicer, Diane Robins, Barbara Wold, Ray Sweet
James Jackson, Israel Lowy, James Michael Roberts
Gek Kee Sim, Saul Silverstein, Richard Axel

When cultured mammalian cells are exposed to DNA, a small subpopulation stably integrate exogenous genes into their chromosomes in a form which is recognized by the replicative and transcriptional apparatus of the host cell. This process is known as transformation (*1*). The transforming elements can be maintained within the host genome for hundreds of generations and frequently express products which alter the phenotypes of the recipient cell. Since transformation in most cell populations is a rare event, identification of transformants requires the use of genes coding for either selectable or readily identifiable functions. Thus, DNA from viruses or eukaryotic cells has been used to transfer genes coding for growth transformation (*2*), thymidine kinase (TK) (*3-7*), adenine phosphoribosyltransferase (APRT) (*8*), and hypoxanthine-guanine phosphoribosyltransferase (HGPRT) (*9, 10*) to mutant cells deficient in these functions.

Transformation therefore provides an opportunity to alter the genotype of a cell by the stable introduction of new genetic information and to examine the expression of exogenous DNA sequences in the transformed host. We will discuss the application of transformation to four basic areas of eukaryotic genetics. (i) The integration of transforming elements into the chromosome, as well as their excision from the chromosome, may involve recombinational systems which reflect the capacity of a somatic cell to reorganize its genome. (ii) The ability to introduce specific wild-type and mutant genes into new cellular environments provides a system in which the functional significance of various features of DNA sequence organization can be studied in vivo. (iii) Transformation has facilitated the isolation of cellular genes coding for APRT and TK; for these two genes classical methods of molecular cloning dependent on messenger RNA (mRNA) enrichment are exceedingly difficult (*11, 12*). (iv) Transformation can be used to analyze the molecular nature of mutation and phenotypic variation in somatic cells.

Dr. Pellicer is an assistant professor of pathology at New York University, New York 10012. Drs. Robins, Wold, Jackson, and Sim are postdoctoral fellows and I. Lowy and J. M. Roberts are graduate students at the College of Physicians & Surgeons, Columbia University, New York 10032. Dr. Sweet is a senior staff associate at the Institute of Cancer Research, Dr. Silverstein is an associate professor of microbiology, and Dr. Axel is a professor of biochemistry and pathology at the College of Physicians & Surgeons.

Viral Thymidine Kinase as a Model System

The development of a successful transformation system for the transfer of eukaryotic genes was initially dependent on the appropriate choice of three basic components: a source of DNA coding for a readily selectable biochemical function, a competent recipient cell deficient in this function, and a selection schema permitting the identification of the rare transformant. In our initial studies, we developed a model system to effect the isolation and transfer of a specific DNA fragment containing the thymidine kinase gene from the herpes simplex virus (HSV-1) genome (*4*). The choice of this system was dictated by several consid-

0036-8075/80/0919-1414$02.00/0