MOLECULAR AND CELLULAR BIOLOGY, Aug. 1986, p. 2784–2791
0270-7306/86/082784-08$02.00/0

Vol. 6, No. 8

Control of Transcription Initiation In Vitro Requires Binding of a Transcription Factor to the Distal Promoter of the Ovalbumin Gene

MARTINE PASTORCIC, HENG WANG, ALEX ELBRECHT, SOPHIA Y. TSAI, MING-JER TSAI, AND BERT W. O'MALLEY*

Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030

Received 16 October 1985/Accepted 22 April 1986

We used a cell-free HeLa cell transcription system to identify and characterize transcription factors and the promoter elements that they recognize in RNA polymerase II-transcribed genes. Deletion of the region (−71 to −83) containing the GTCAAA direct repeat resulted in a marked decrease of specific transcription of the ovalbumin gene; transcription could be competed with DNA fragments containing this sequence. Furthermore, DNase I footprinting identified a protein-binding site including this direct repeat with crude extracts and one of the partially purified protein fractions required for transcription. We propose that a soluble factor activates transcription through binding to the direct repeat of GTCAAA sequence upstream from the ovalbumin gene.

Recombinant DNA methodology and the techniques of gene transfer have facilitated the characterization of promoters for RNA polymerase II-transcribed genes. The efficiency of mutated promoters in directing accurate initiation has been tested in vivo and in vitro (1, 2, 10, 32). These experiments suggested the existence of distinct promoter elements upstream from the initiation site of transcription. The TATA box, a highly conserved homology among mRNA coding genes and located at −24 to −32 base pairs from the initiation site, is essential for efficient and accurate initiation of most genes both in vivo and in vitro (2, 32). In addition, sequences upstream from the TATA box have been shown to be required in vivo (32) and in vitro (20, 22, 24, 27, 29, 40, 41) for efficient transcription of many eucaryotic genes. This is clearly the case for the ovalbumin gene in vivo (28). We previously showed that accurate initiation of the ovalbumin gene transcription in vitro requires the TATA homology (39, 42). We demonstrate here that sequences upstream from the TATA box are important for quantitative transcription of the ovalbumin gene in crude extracts from HeLa cells. This upstream promoter element is located approximately 80 base pairs from the cap site and includes a direct repeat of GTCAAA sequence. Since this region was originally defined by Benoist et al. (1) and Efstratiadis et al. (11) as the CAAT-box region, we tentatively designate the direct repeat sequence as the CAAT box for reference purposes.

The requirement for a distal promoter element was initially indicated by 5′ deletion mapping. To test whether a transcription factor interacts with this promoter element, we performed transcription competition assays and found that crucial sequences for the binding of a transcription factor include the CAAT box. We have shown previously by exonuclease footprinting that a protein which binds to this region of the ovalbumin gene is present in HeLa cell fractions (13). We now confirm and extend this footprinting analysis using DNase I with either crude nuclear extracts or the partially purified fractions (38) required for transcription in reconstituted systems. These results suggest that the upstream element delineated by the functional transcription assays is the specific binding site for a positive effector of transcription.

* Corresponding author.

MATERIALS AND METHODS

Plasmids and DNA fragments. The construction of pSV.OG (ovalbumin-globin fusion gene or ovalglobin gene) and the corresponding 5′ deletions of the 5′-flanking region of the ovalbumin gene were described previously (28) and are summarized in Fig. 1.

The plasmid templates for the synthesis of RNA probes by SP6 RNA polymerase were constructed by inserting ovalglobin or simian virus 40 (SV40) DNA fragments into the polylinker of pSP65 or pSP64, respectively (Promega Biotech, Madison, Wis.) as described in the legend to Fig. 1. The ovalbumin RNA probe used in the competition assays was synthesized from pSPOV0.4 in which ovalbumin sequence from the *Ava*II site to the *Hin*dIII site (position −135 to +236) was cloned into the corresponding *Hin*dIII site and *Bam*HI site of the pSP64 polylinker. The template was linearized with *Bam*HI.

DNA fragments used in the competition experiments were isolated from 3′ deletion mutants constructed in our laboratory by L. Lopez (unpublished data) using BAL 31 digestion. The endpoints of the deletions were determined by Maxam and Gilbert sequencing (34). DNA competitors were *Hae*III fragments purified on 4% polyacrylamide gels and isolated as described by Maniatis et al. (31). They bear ovalbumin upstream sequence from position −753 to the endpoint of the 3′ deletion, the 8-base-pair *Cla*I linker, and 41 base pairs of pBR322 (from the *Cla*I site at position 23 to the *Hae*III site at position 4343).

Preparation and fractionation of crude extracts. Nuclear extracts were prepared by the procedure of Dignam et al. (7), except that preparations directly used in transcription assays were concentrated by dialysis against 10 volumes of 20 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid) (pH 7.9)–100 mM KCl–0.2 mM EDTA–0.5 mM dithiothreitol (DTT)–0.5 mM phenylmethylsulfonyl fluoride–20% (vol/vol) glycerol containing 30% (wt/vol) polyethylene glycol 20000 until the protein concentration reached 15 to 20 mg/ml; these preparations were further dialyzed against 50 volumes of the same buffer without polyethylene glycol for an additional 5 to 8 h. Whole-cell extracts were prepared by a modified procedure of Manley et al. (33) as described by Tsai et al. (38). Both nuclear extracts and whole-cell extracts were stored at −84°C.

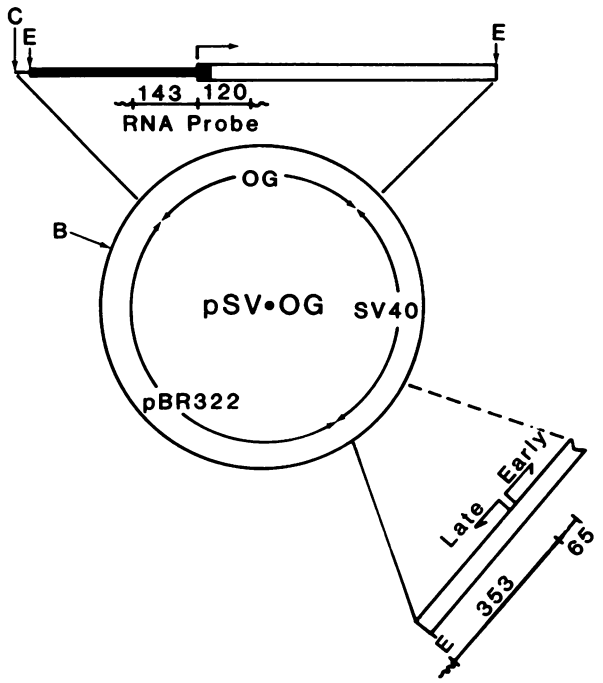

FIG. 1. Structure of pSV.OG plasmid and RNA probes used in the analysis of ovalglobin and SV40 early gene transcription. The construction of pSV.OG was described previously (28). In the ovalglobin (OG) plasmid, ovalbumin sequences (solid box) from position −753 to position +41 are fused to adult chicken β-globin sequences (open box). The start and direction of ovalglobin and SV40 early and late transcription are indicated by arrows in the expanded regions. 5′ deletions of the flanking sequences of the ovalbumin gene were derived from pSV.OG by BAL 31 digestion from the *Cla*I site of pBR322 as described previously (28). The template for the synthesis of the ovalglobin RNA probe was constructed by inserting the *Cla*I (position −143) to *Pvu*II (position +120) fragment of the 5′ deletion mutant pSV.OG(−143) between the *Acc*I and *Sma*I sites of pSP65. The RNA probe was synthesized from this plasmid after being linearized with *Hin*dIII. It consists of 263 nucleotides of antisense ovalbumin gene sequences and 23 nucleotides of vector sequences. The *Hpa*II (position 347) to *Hin*dIII (position 5172) sequence of the SV40 early region was inserted between the *Eco*RI and *Hin*dIII sites of pSP64. The probe synthesized from this plasmid consists of 418 nucleotides of antisense SV40 sequences and 6 nucleotides of vector. E, *Eco*RI; C, *Cla*I; B, *Bam*HI.

Fractionation of nuclear extracts was essentially as described previously with whole-cell extracts (38). Briefly, the extract was first loaded on a DEAE-cellulose column in 50 mM $(NH_4)_2SO_4$, and the bound proteins were eluted stepwise with 175 mM (DE175 fraction) and 500 mM (DE500 fraction) $(NH_4)_2SO_4$. The unbound fraction was further fractionated on a phosphocellulose column. The loading was at 100 mM KCl, and the bound proteins were eluted in two steps with 350 mM (P350 fraction) and 1 M (P1000 fraction) KCl, respectively.

In vitro transcription and RNA isolation. Standard 100-μl reaction mixtures contained 12 mM HEPES (pH 7.9), 3.75 mM $MgCl_2$, 60 mM KCl, 1 mM DTT, 0.1 mM EDTA, 12% glycerol, and 500 mM each ATP, CTP, UTP, and GTP. The mixture contained 1 to 2 μg of DNA and 30 μl of nuclear extract or 60 μl of whole-cell extract. The DNA templates were either supercoiled or linearized plasmids or purified DNA fragments. Incubations were at 30°C for 40 min. At completion of the incubation, 400 μl of 50 mM Tris-hydrochloride (pH 7.5)–10 mM EDTA–0.5% sodium dodecyl sulfate (SDS) containing 200 μg of proteinase K and 10 μg of yeast tRNA was added and then incubated at 37°C for 30 min. To isolate RNA, samples (500 μl) treated with proteinase K were added to 1 ml of 50 mM Tris-hydrochloride (pH 7.5)–25 mM EDTA–1% Sarkosyl–1 g of CsCl and gently overlayed onto 1.5 ml of 5.7 M CsCl–100 mM EDTA (pH 7.4). The RNA was pelleted at 35,000 rpm at 16°C for 16 h in a Beckman SW50.1 rotor (20). The layers of the supernatant fraction were carefully removed, and the RNA pellets were suspended in 0.1% SDS and precipitated with ethanol. After a second ethanol precipitation in 0.3 M sodium acetate, the RNA samples were analyzed by RNase mapping.

Synthesis of SP6 probes. Transcription by SP6 polymerase was essentially as described by Green et al. (19) and Zinn et al. (43). Reaction mixtures (40 μl) contained 3 μg of linear template, 40 mM Tris hydrochloride (pH 7.5), 6 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 500 μM each ATP, UTP, and CTP, 50 μM GTP, 60 μCi of [α-^{32}P]GTP (410 Ci/mmol; Amersham Corp., Arlington Heights, Ill.), and 15 U of SP6 RNA polymerase (Promega). Incubations were at 37°C for 1 h. RNase-free DNase I was then added to 20 μg/ml, and the reactions were incubated further at room temperature for 20 min. After the addition of 150 μl of 0.1% SDS–10 mM EDTA–7 M urea–10 μg of tRNA, the RNA was extracted twice with phenol-chloroform-isoamyl alcohol (25:24:1, vol/vol/vol) and precipitated twice with 1 M ammonium acetate.

RNase analysis of RNA transcripts. Hybridization of RNA probes to transcription products was carried out as follows. Dried RNA samples were dissolved in 30 μl of hybridization buffer containing 60% formamide, 10 mM HEPES (pH 7.5), 2 mM EDTA, 600 mM NaCl, and 2×10^5 to 5×10^5 cpm of RNA probe, heated at 100°C for 5 min, and incubated at 54°C for 10 to 14 h. A 370-μl portion of 10 mM Tris-hydrochloride (pH 7.5)–5 mM EDTA–300 mM NaCl containing 30 μg of RNase A was added to remove unhybridized single-strand RNA probe at 24°C for 30 min. The reactions were stopped by the addition of 20 μl of 10% SDS and digestion with 40 μg of proteinase K for 30 min at 24°C. After incubation, 10 μg of tRNA was added, and the reaction mixtures were phenol extracted and ethanol precipitated. The RNA samples were then dissolved in 99% formamide, denatured, and analyzed on 5% polyacrylamide–7 M urea denaturing gels as described previously (38).

Competition assays. Competitor DNA (0 to 3 μg) was incubated for 20 min on ice in 90-μl reaction volumes containing 13 mM HEPES (pH 7.9), 4 mM $MgCl_2$, 66 mM KCl, 0.1 mM EDTA, 1 mM DTT, 12% glycerol, and 20 μl of nuclear extract. The nucleotide triphosphates were then added in 10 μl along with 1 μg of linear DNA template pSV.OG(−161), and the mixture was further incubated at 30°C for 30 min. Reactions were stopped and processed as described in standard assays. The RNA probe used (synthesized from pSPOV0.4) contains ovalbumin sequences from −135 to +226. The reaction products were analyzed on 12% polyacrylamide–7 M urea denaturing gels.

^{32}P labeling of DNA fragments. For DNase I footprinting, pOV0.77 was digested with *Cla*I and endlabeled with [α-^{32}P]dCTP (3,000 Ci/mmol; ICN Pharmaceuticals Inc., Irvine, Calif.) with the Klenow fragment of DNA polymerase I as described by Maniatis et al. (31). A second digestion with *Sau*3A produced a 368-base-pair DNA fragment 3′ labeled on the noncoding strand which was purified by polyacrylamide gel electrophoresis. DNase I footprints on the coding strand were carried out with a 281-base-pair *Cla*I to *Pvu*II

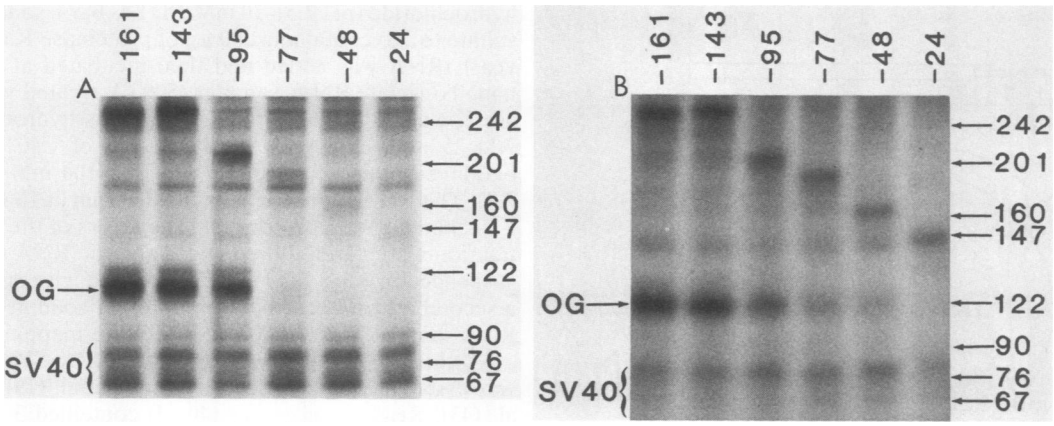

FIG. 2. In vitro transcription of pSV.OG and its 5′ deletion mutants. RNA was synthesized in a 100-μl reaction mixture with 2 μg of circular DNA template and 30 μl of nuclear extract (A) or 60 μl of whole-cell extract (B). Each template contains a deletion of the 5′-flanking sequences of the ovalbumin gene promoter. The endpoints of the deletion are indicated above the corresponding lanes. RNA products were then analyzed for the ovalglobin (OG) and SV40 early transcripts by an RNase A mapping technique. The numbers on the right indicate the position of DNA size standards (in nucleotides). Transcription was quantified by densitometric scanning of low-exposure autoradiograms.

fragment of pSV.OG(−161) 3′ labeled at the *Cla*I site, as described above. The typical specific activity was 2×10^7 cpm/μg of DNA.

DNase I footprinting reactions. DNase I footprinting was carried out as described by Galas and Schmitz (16) and modified by Parker and Topol (37). Reaction mixtures (20 μl) containing 18 mM HEPES (pH 7.9), 90 mM KCl, 4.5 mM $MgCl_2$, 0.2 mM EDTA, 1.8 mM DTT, 18% glycerol, 400 ng of pBR322, and 7 ng of ^{32}P-labeled *Sau*3A-*Cla*I or *Cla*I-*Pvu*II fragment were incubated with either HeLa cell nuclear extract or the P1000 fraction at room temperature for 5 min. Pancreatic DNase I (Worthington Diagnostics, Freehold, N.J.) was then added and incubated for 30 s at room temperature. The reaction was stopped by the addition of 100 μl of 10 mM EDTA–0.1% SDS–50 μg of proteinase K per ml–1.5 μg of pBR322 and incubated at 37°C for 15 min. After 3 min at 90°C the samples were extracted with phenol-chloroform, ethanol precipitated, and loaded on 8% polyacrylamide–7 M urea sequencing gels.

RESULTS

Effect of 5′ deletions on transcription of the ovalglobin gene. We used 5′ deletions of the ovalglobin gene (Fig. 1) in an in vitro transcription assay and an RNase A mapping technique (19) to search for DNA sequences upstream from the TATA box that affect transcription. Since the ^{32}P-labeled RNA probe contains sequences from −143 to +120, accurate initiation results in a transcript which protects 120 nucleotides of probe (Fig. 2A). We found that deletions of sequences upstream from −95 have little effect on promoter strength. Transcription was reduced by about 20% by deletion of sequences between −143 and −95. However, when the upstream sequences were deleted to −77, the transcriptional activity of the template was reduced further by more than 80% (Fig. 2A). The residual level of transcription was abolished when the TATA box was deleted. Since the intensity of the signal at 120 nucleotides is proportional to the amount of specific transcript present (data not shown), we conclude that the 5′-flanking sequence between −77 and −95 is required for efficient and accurate initiation. The readthrough signal at 260 nucleotides observed with pSV.OG(−161) or pSV.OG(−143) is due either to transcription initiated on the sequences upstream from −143 or to specific initiation events that do not terminate after one turn on the plasmid. The latter is more likely since the amount of readthrough transcripts decreased along with the amount of specific transcription when upstream sequences were deleted.

As an internal reference we also probed the RNA products of the transcription assay with sequences complementary to SV40 early transcripts. Heterogeneous initiation of SV40 early transcripts produced a group of protected fragments, the smallest of which was approximately 65 nucleotides long. These results show that deletion of sequences in the ovalbumin promoter does not affect transcription of the SV40 sequences in the plasmid. Thus, this assay specifically measures ovalbumin promoter activity in vitro.

To eliminate the possibility that the presence of the SV40 early and late promoters might affect the results obtained, we isolated the *Bam*HI-*Eco*RI fragments which contain only the ovalglobin gene and its promoter and used them in the in vitro transcription system. Again the upstream sequence dependence of the ovalbumin promoter was observed, the results being essentially identical to those presented in Fig. 2A (data not shown).

The same sequences affected transcription with whole-cell extracts (Fig. 2B). In this case the major decrease in transcription efficiency after the deletion of sequences between −95 and −77 was slightly lower (about 60%) than that obtained with nuclear extracts. On the other hand, the decrease in transcription resulting from deletion of sequences between −143 and −95 was slightly higher (about 40%). Deletion of sequences from −77 to −48 resulted in a further decrease in transcription as observed with nuclear extracts. Despite these minor differences, a major decrease in transcription followed deletion of sequences from −77 to −95 with both extracts.

We also tested the effect of various assay conditions on the upstream sequence dependence of transcription. Template DNA concentration (from 5 to 50 μg/ml), $MgCl_2$ concentration (from 2.5 to 9 mM), NaCl concentration (from 60 to 100 mM), and amount of nuclear extract (from 2 to 10 mg of protein per ml) did not appreciably alter the results described above (data not shown).

Following our 5′ deletion mapping, an upstream promoter can be operationally defined as the interval of deletion that affects transcription most. Since this interval (−95 to −77)

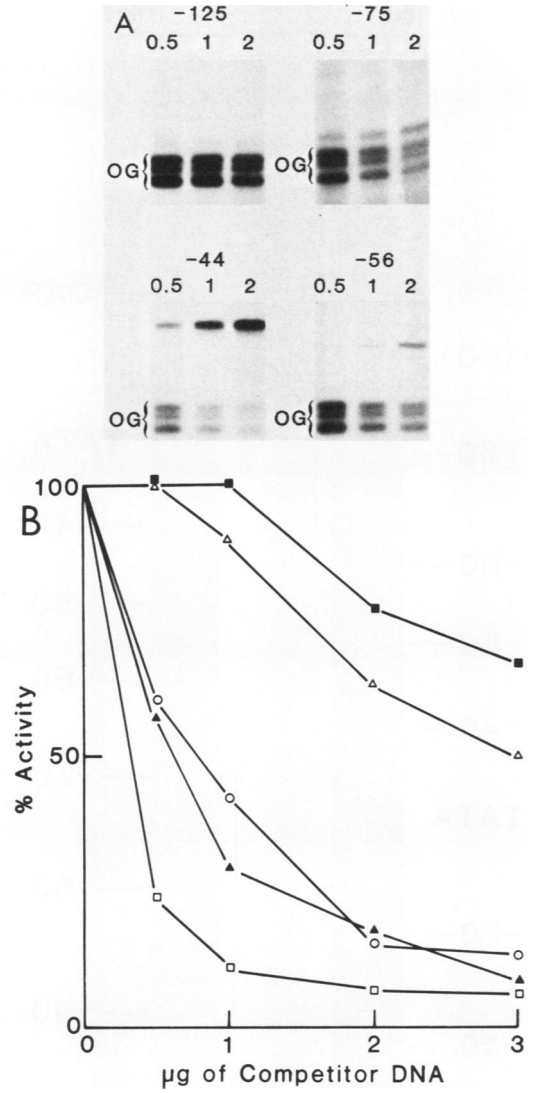

FIG. 3. Competition of in vitro transcription by ovalbumin gene upstream sequences. (A) Transcription competition assays and the DNA fragments used as competitors are described in the text. The 3′ ends of the ovalbumin flanking sequences present on each competitor (−125, −75, −56, and −44) and the amount of competitor DNA (0.5, 1, and 2 μg) are indicated. (B) Transcription of the ovalbumin gene in the presence of increasing amounts of competitors was estimated by densitometric scanning of low-exposure autoradiograms of gels such as those displayed in panel A. 100% activity was set as the amount of transcript obtained upon the addition of 0.5 μg of −125 competitor DNA. Competitor DNAs used were −125 (■), −89 (△), −75 (○), −56 (▲) and −44 (□).

contains one of the two direct repeats of the CAAT-box consensus sequence, we refer to it as the CAAT-box region (see Fig. 7 for sequences).

Competition of in vitro transcription by ovalbumin sequences. To assess whether the *cis*-acting element delineated by 5′ deletion mapping is a binding site for a transcription factor, we carried out competition assays. Competitors were blunt-ended DNA fragments of ovalbumin upstream sequences with various 3′ ends that did not contain the TATA box. They included ovalbumin sequences from position −753 to the end of the 3′ deletion, linked to 41 base pairs of pBR322 DNA. Since the RNA probe used in this assay contained only ovalbumin sequences, specific transcripts should protect a 41-nucleotide fragment. Three major RNA products were observed, differing in size by 2 to 6 nucleotides. We do not know whether these differences are due to heterogeneous transcription starts observed in vivo (30) or to residual unpaired residues resistant to RNase A digestion.

DNA fragments with 3′ endpoints at −125 and −89 were ineffective in reducing transcription of the template pSV.OG(−161) even at a molar ratio of competitor to template of 5 or 10 (0.5 and 1 μg; Fig. 3A and B). In contrast, fragments with 3′ endpoints at −75 or −56 reduced transcription to approximately 60% at a fivefold molar excess and to 30% at a 10-fold molar excess. These results suggest that the CAAT-box region might be a binding site for a transcription factor and that sequences which are critical for stable binding of the protein are localized between −89 and −75. It should be pointed out that the most effective competitor was a fragment containing ovalbumin sequences further downstream to −44 (Fig. 3A and B). Thus, sequences between −56 and −44 could influence the binding of the CAAT-box factor or perhaps may even participate in the binding of another transcription factor.

As a control, using the same assay conditions we also monitored transcription from the SV40 early promoter which lacks a CAAT box (Fig. 4A). Competitor DNA fragments with or without the CAAT box had no differential effect on transcription of the SV40 early promoter. Furthermore, the basal level of transcription from an ovalbumin DNA template in which the CAAT box is deleted [pSV.OG(−48); Fig. 4B] was similar whether the competitor DNA contained the CAAT box or not. This suggests that the presence of the CAAT box on DNA fragments does not result in an increased ability to compete for a transcription factor generally required for the transcription of SV40 or for the basal level of ovalbumin gene expression on pSV.OG(−48).

Footprinting. If the CAAT-box region is the binding site for a transcription factor, it should be possible to visualize it by DNase I footprinting. This would also provide a convenient assay in the purification of the protein. Using crude nuclear extracts from HeLa cells, we were able to obtain a footprint on both the noncoding and coding strands (Fig. 5). The binding activity protects a region from approximately

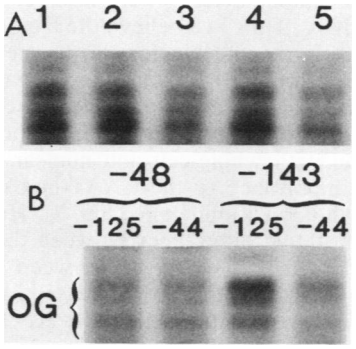

FIG. 4. Competition of transcription of SV40 and a CAAT-box-minus ovalbumin template. (A) Assays were as described in the legend to Fig. 3A, except that SV40 transcription was monitored. Lanes: 1, no competitor added; 2 and 3, 1 and 3 μg, respectively, of ovalbumin sequence fragment with 3′ end at −125; 4 and 5, 1 and 3 μg, respectively, of fragment with 3′ end at −44. (B) Competition assays were carried out as described in the text, except that the templates [pSV.OG(−143) and pSV.OG(−48)] were circular DNA. Competitor (−125 or −44) (3 μg) was added with 60 μl of whole-cell extract in 100-μl reaction mixtures. OG, Ovalglobin.

−70 to −90. Since a crude nuclear extract was used, it is not surprising that additional minor binding sites are observed between −60 to −40 and around the TATA box.

When we fractionated the crude nuclear extract on DEAE-Sephadex and phosphocellulose columns by the method of Tsai et al. (38), we found that most of the CAAT-box binding activity was localized in the P1000 fraction (Fig. 6; data not shown). In addition, the results in Fig. 6 show that the amount of binding in the CAAT-box region is proportional to the amount of P1000 added. In this case the binding site extends from −66 to −89 with sites at −66 and −68 showing moderately enhanced cleavage by DNase I. Similar hypersensitive sites have been reported by others using the DNase I footprinting method (15, 23).

DISCUSSION

The results described in this paper suggest that a distal promoter element identified as the CAAT-box region regulates transcription initiation in vitro by binding a transcription factor. To support such a mechanism, we present data from both in vitro transcription assays and DNase I footprinting analysis.

Using 5′ deletions of an ovalbumin gene template in a cell-free HeLa cell transcription system, we showed that deletion of the CAAT-box region results in a marked decrease in transcription. Transcriptional activity was competed specifically by DNA fragments containing the CAAT-box region, suggesting that such activity requires the presence of a *trans*-acting factor in the HeLa cell nuclear extract. Furthermore, evidence for a protein binding to the CAAT box was obtained by DNase I footprinting.

We fractionated the nuclear extracts according to a scheme, previously described for whole-cell extracts, with DEAE-Sephadex and phosphocellulose columns (38). In a reconstituted system (38) with partially purified fraction instead of crude extracts, transcription efficiency also depends on upstream sequences (M. Pastorcic, unpublished data). One fraction termed P1000 which is absolutely essential for correct initiation (38) also contains most of the CAAT-box-binding activity (13). Recently, we further fractionated the CAAT-box-binding factor by Sephacryl 300 chromatography. We found that those fractions containing the CAAT-box-binding activity were also required for transcriptional activity. Thus, we believe the factor(s) binding to the CAAT box is responsible for the upstream-sequence-dependent transcription of the ovalbumin gene. Final identification of the binding activity as a transcription factor awaits purification of the protein to homogeneity.

Transcription with whole-cell and nuclear extracts both indicated the importance of the CAAT-box region for the expression of the ovalbumin gene (Fig. 2). However, there are some quantitative differences between the two assays; for example, the effect of sequences between −143 and −95 is more apparent in whole-cell extracts than in nuclear extracts. Similarly, the data obtained by Knoll et al. (28) in transfections of HeLa cells with the same plasmids also indicate a 20 to 40% decrease in ovalbumin gene expression upon deletion of sequences between −143 and −95. The reason for these quantitative differences is not known. However, supposing that the interaction of the CAAT-box transcription factor with the −77 to −95 region is influenced by the sequence upstream from −95, these differences might be explained if the concentration of CAAT box transcription factor is high enough to obscure the effect of sequences upstream from −95 in nuclear extracts but not in whole-cell extracts or in vivo. This hypothesis is consistent with the overall reduction of transcription after deletion of sequences between −143 and −77 being greater with nuclear extracts than with whole-cell extracts (Fig. 2).

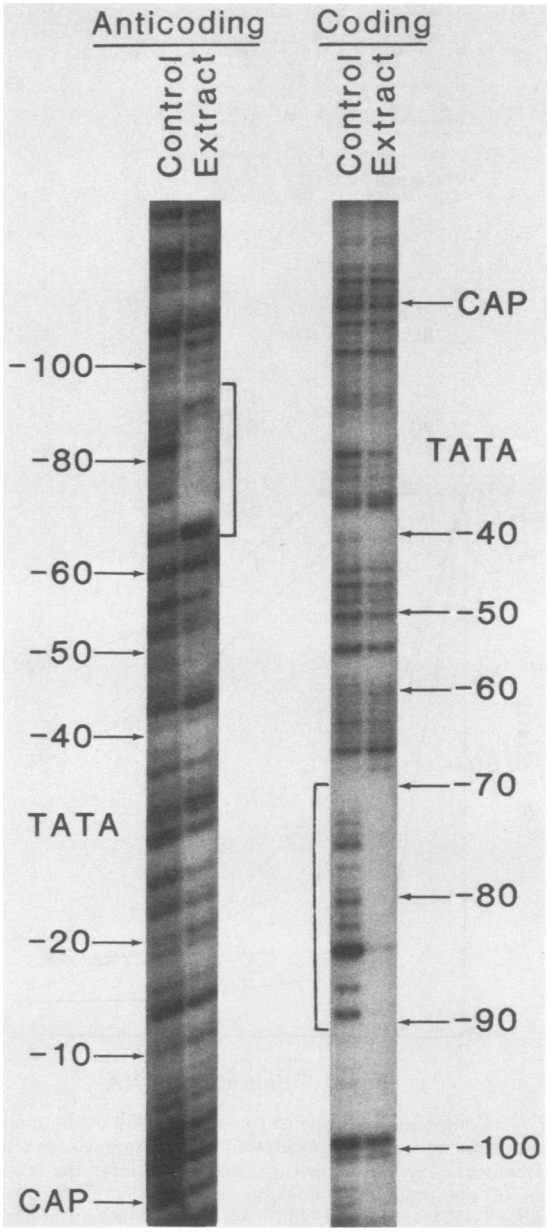

FIG. 5. DNase I footprint of the ovalbumin gene promoter in the presence of nuclear extract. Reactions were carried out as described in the text. Nuclear extract (3 μl) and 0.46 μg (extract lanes) or 0.18 μg (control lanes) of DNase I were used in these experiments. The corresponding positions on the ovalbumin upstream sequences are determined by Maxam and Gilbert sequencing ladders.

Using footprinting with exonuclease, we have demonstrated previously the existence of CAAT-box-binding activity in the P1000 fraction prepared from HeLa whole-cell extracts (13). This binding activity is also present in the P1000 fraction when P1000 is prepared from isolated nuclei. Exonuclease footprints of the nuclear P1000 fraction and the whole-cell P1000 fraction identify the same borders for the CAAT-box binder (from −69 to −90). The DNase I footprints support the data obtained by exonuclease footprint-

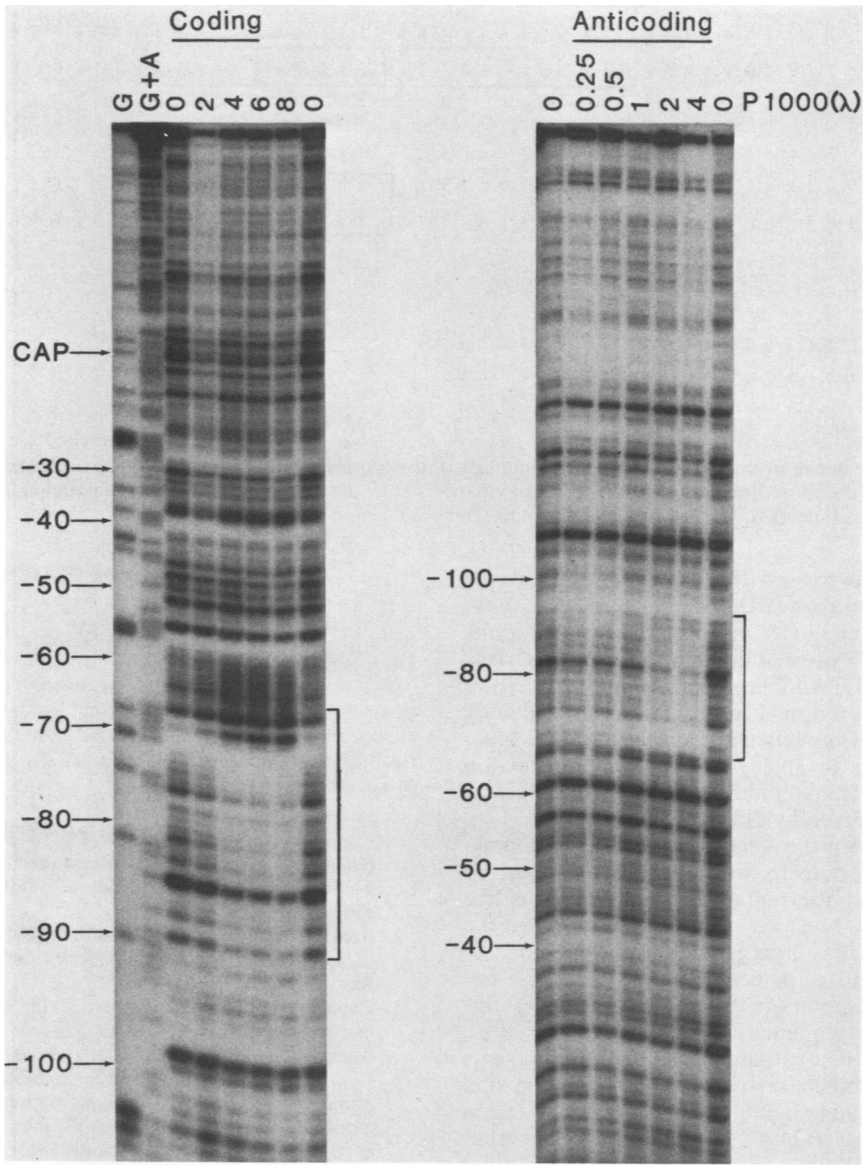

FIG. 6. Protection of the ovalbumin gene promoter from DNase I digestion in the presence of the P1000 fraction. DNase I footprinting reactions on the coding and noncoding strands of the ovalbumin gene promoter were carried out in the presence of increasing amounts of P1000 as indicated. DNase I (0.18 and 0.23 μg) was used in control and P1000 lanes, respectively. G and G+A tracks indicate Maxam and Gilbert sequencing ladders, and the arrows mark the corresponding positions on the ovalbumin gene upstream sequence.

ing. The CAAT-box-binding protein protects a region of approximately 20 base pairs on both the noncoding and coding strands of the DNA. The DNase I footprints also show that other regions between the CAAT box and the transcription initiation site, including the TATA box, are protected from digestion with DNase I. This could suggest that the HeLa nuclear extract contains other specific DNA-binding proteins. Some evidence that such factors could be involved in the regulation of transcription comes from the in vitro transcription competition experiments. The most effective competitor DNA fragment was one that not only contains the CAAT box region but also approximately 35 base pairs of DNA 3′ to the CAAT box.

The region protected in the footprints contains a direct repeat (Fig. 7) homologous to the CAAT-box consensus sequence (1, 11). The importance of the CAAT box for transcription has been suggested by the study of localized mutations in transcription assays in vivo (4, 12, 21, 35). In particular, from studies with point mutations of the rabbit β-globin gene CAAT box, Dierks et al. (6) concluded that any transition in the sequence CCA results in a four- to fivefold reduction in transcription efficiency. Gelinas et al. (17) and Collins et al. (5) have independently identified a natural point mutation in the distal CAAT box of the human $^{A}\gamma$-globin gene. The mutation is associated with hereditary persistence of fetal hemoglobin suggesting that this mutation releases the gene from its normal developmental regulation. These studies indicate that the CAAT-box sequence is important for the expression of eucaryotic genes.

Recently, Jones et al. (26) have identified a protein-binding site including the two CAAT boxes present in reverse orientation upstream from the thymidine kinase gene. Linker scanner mutations affecting the herpes simplex virus distal CAAT box reduce transcription in vivo (14, 35) and in vitro

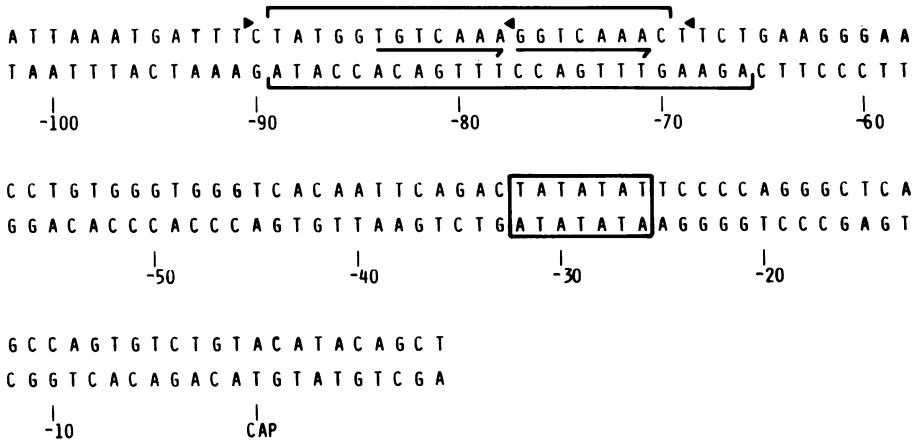

FIG. 7. DNA sequence of the ovalbumin promoter and summary of the footprinting experiments. DNase I footprints on both coding and noncoding strands are indicated by brackets. Arrowheads show barriers to exonuclease digestion on the noncoding strand (13). Horizontal arrows underline the direct repeats of the CAAT-box sequence. The TATA homology is boxed.

(26). These results also suggest that a transcription factor binds to a promoter element which includes the CAAT box. Most recently, Graves et al. (18) introduced point mutations in the sequence CCAAT present in the herpes simplex virus thymidine kinase distal CAAT box and about 80 base pairs upstream from the long terminal repeat cap site of Moloney murine sarcoma virus. Equivalent mutations affect the function of both promoters in analogous ways in transfection assays and also the DNA-binding activity of a nuclear protein fraction, as assayed by DNase I footprinting.

The homology between the −80 region of the ovalbumin gene and the CAAT-consensus was originally reported by Benoist et al. (1), and we tentatively refer to it as the CAAT-box region. However, we wish to emphasize that the GTCAAA direct repeats diverge significantly from the GG$^{C}_{T}$CAAATCT consensus sequence and the CCAAT homologies present in promoters in which the functional reference of the consensus sequence has been documented (6, 18). It is possible that the ovalbumin −80 region represents a novel transcription control region and factor-binding site. It is also possible that the homology between the −80 region and the more conserved CCAAT box could reflect a relatedness between the ovalbumin transcription factor and other CAAT factors as members of a larger family of regulatory proteins. To distinguish these possibilities, we are presently analyzing the effects of competitor DNA fragments from various promoters on the transcription of the ovalbumin gene and the specific binding to the −80 region.

It is now clear from studies with several eucaryotic genes that distal promoter elements are necessary for transcription initiation. The CAAT box sequence itself is not a ubiquitous component of eucaryotic promoters, and a variety of regulatory factors could be necessary to interact with the different upstream elements characterized thus far. Indeed, interactions between specific transcription factors and various promoter elements have been described for heat shock genes (37), histone genes (3), SV40 genes (8, 9), yeast mating-type genes (25), and the alcohol dehydrogenase gene (23). However, since we have identified a CAAT-box binder in a heterologous system, it is suggested that this protein is conserved among species.

To help answer these questions and to determine its molecular complexity, we are purifying the CAAT-box binder and studying its function in transcription assays in vitro with different templates.

ACKNOWLEDGMENTS

We wish to thank Geoffrey Long and Peter Hine for excellent technical assistance and Linda Glazener and Wanda Beattie for DNA sequencing.

A.E. acknowledges the support of the Medical Research Council of Canada in the form of a postdoctoral fellowship. This work was supported by Public Health Service grants HD-08188 and HD-07495 (to B.W.O.) and HD-17379 (to M.J.T.) from the National Institutes of Health.

LITERATURE CITED

1. **Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon.** 1980. The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res. **8**:127–142.
2. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
3. **Capasso, O., and N. Heintz.** 1985. Regulated expression of mammalian histone H4 genes *in vivo* requires a trans-acting transcription factor. Proc. Natl. Acad. Sci. USA **82**:5622–5626.
4. **Charnay, P., P. Mellon, and T. Maniatis.** 1985. Linker scanning mutagenesis of the 5′-flanking region of the mouse β-major globin gene: sequence requirements for transcription in erythroid and nonerythroid cells. Mol. Cell. Biol. **5**:1498–1511.
5. **Collins, F. S., J. E. Metherall, M. Yamakawa, J. Pan, S. M. Weissman, and B. Forget.** 1985. A point mutation of the $^{A}\gamma$-globin gene promoter in Greek hereditary persistence of fetal haemoglobin. Nature (London) **313**:325–326.
6. **Dierks, P., A. van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissmann.** 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell **32**:695–706.
7. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11**:1475–1489.
8. **Dynan, W. S., and R. Tjian.** 1983. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell **32**:669–680.
9. **Dynan, W. S., and R. Tjian.** 1983. A promoter specific transcription factor sp1 binds to upstream sequences in the SV40 early promoter. Cell **35**:79–82.
10. **Dynan, W. S., and R. Tjian.** 1985. Control of eucaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature (London) **316**:774–778.
11. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the

β-globin gene family. Cell **21**:653–668.
12. **Eisenberg, S. P., D. M. Coen, and S. L. McKnight.** 1985. Promoter domains required for expression of plasmid-borne copies of the herpes simplex virus thymidine kinase gene in virus-infected mouse fibroblasts and microinjected frog oocytes. Mol. Cell. Biol. **5**:1940–1947.
13. **Elbrecht, A., S. Y. Tsai, M.-J. Tsai, and B. W. O'Malley.** 1985. Identification by exonuclease footprinting of a distal promoter-binding protein from HeLa cell extracts. DNA **4**:233–240.
14. **ElKareh, A., A. J. M. Murphy, T. Fichter, A. Efstratiadis, and S. Silverstein.** 1985. "Transactivation" control signals in the promoter of the herpes virus thymidine kinase gene. Proc. Natl. Acad. Sci. USA **82**:1002–1006.
15. **Emerson, B. M., C. D. Lewis, and G. Felsenfeld.** 1985. Interaction of specific nuclear factors with the nuclease-hypersensitive region of the chicken adult β-globin gene: nature of the binding domain. Cell **41**:21–30.
16. **Galas, D., and A. Schmitz.** 1978. DNase I footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. **5**:3157–3170.
17. **Gelinas, R., B. Endlich, C. Pfeiffer, M. Yagi, and G. Stamatoyannopoulos.** 1985. G to A substitution in the distal CCAAT box of the $^{A}\gamma$-globin gene in Greek hereditary persistence of fetal haemoglobin. Nature (London) **313**:323–324.
18. **Graves, B. J., P. F. Johnson, and S. L. McKnight.** 1986. Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. Cell **44**:565–576.
19. **Green, M. R., T. Maniatis, and D. A. Melton.** 1982. Human β-globin pre-mRNA synthesized *in vitro* is accurately spliced in *Xenopus* oocyte nuclei. Cell **32**:681–694.
20. **Grosschedl, R., and M. L. Birnstiel.** 1982. Delineation of far upstream sequences required for maximal *in vitro* transcription of an H2A histone gene. Proc. Natl. Acad. Sci. USA **79**:297–301.
21. **Grosveld, G. C., A. Rosenthal, and R. A. Flavell.** 1982. Sequence requirements for the transcription of the rabbit β-globin gene *in vivo*: the −80 region. Nucleic Acids Res. **10**:4951–4971.
22. **Hansen, U., and P. A. Sharp.** 1983. Sequences controlling *in vitro* transcription of SV40 promoters. EMBO J. **2**:2293–2303.
23. **Heberlein, U., B. England, and R. Tjian.** 1985. Characterization of *Drosophila* transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. Cell **41**:965–977.
24. **Hen, P., P. Sassone-Corsi, J. Corden, M. P. Gaub, and P. Chambon.** 1982. Sequences upstream from the TATA box are required *in vivo* and *in vitro* for efficient transcription from the adenovirus serotype 2 major late promoter. Proc. Natl. Acad. Sci. USA **79**:7132–7136.
25. **Johnson, A. D., and I. Herskowitz.** 1985. A repressor (MAT α2 product) and its operator control expression of a set of cell type specific genes in yeast. Cell **42**:237–247.
26. **Jones, K. A., K. R. Yamamoto, and R. Tjian.** 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter *in vitro*. Cell **42**:559–572.
27. **Jove, R., and J. L. Manley.** 1984. *In vitro* transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter proximal sites. J. Biol. Chem. **259**:8513–8521.
28. **Knoll, B. J., T. Zarucki-Schulz, D. C. Dean, and B. W. O'Malley.** 1983. Definition of the ovalbumin gene promoter by transfer of an ovalglobin fusion gene into cultured cells. Nucleic Acids Res. **11**:6733–6754.
29. **Lebowitz, P., and P. K. Ghosh.** 1982. Initiation and regulation of simian virus 40 early transcription in vitro. J. Virol. **41**:449–461.
30. **Malek, L. T., W. H. Eschenfeldt, T. W. Munns, and R. E. Rhoads.** 1981. Heterogeneity of the 5′-terminus of the hen ovalbumin messenger ribonucleic acid. Nucleic Acids Res. **9**:1657–1673.
31. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Manley, J. L.** 1983. Analysis of the expression of genes encoding animal mRNA by *in vitro* techniques. Prog. Nucleic Acids Res. Mol. Biol. **30**:195–244.
33. **Manley, J. L., A. Fire, A. Cano, P. A. Sharp, and M. L. Gefter.** 1980. DNA-dependent transcription of adenovirus genes in a soluble whole cell extract. Proc. Natl. Acad. Sci. USA **77**:3855–3859.
34. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
35. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of a eucaryotic protein-coding gene. Science **217**:316–324.
36. **Myers, R. M., D. C. Rio, A. K. Robbins, and R. Tjian.** 1981. SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. Cell **25**:373–384.
37. **Parker, C. S., and J. Topol.** 1984. A *Drosophila* RNA polymerase II transcription factor for the heat-shock gene binds to the regulatory site of an hsp 70 gene. Cell **37**:273–283.
38. **Tsai, S. Y., M.-J. Tsai, L. E. Kops, P. P. Minghetti, and B. W. O'Malley.** 1981. Transcription factors from oviduct and HeLa cells are similar. J. Biol. Chem. **255**:13055–13059.
39. **Tsai, S. Y., M.-J. Tsai, and B. W. O'Malley.** 1981. Specific 5′-flanking sequences are required for faithful initiation of *in vitro* transcription of the ovalbumin gene. Proc. Natl. Acad. Sci. USA **78**:879–883.
40. **Tsuda, M., and Y. Suzuki.** 1981. Faithful transcription initiation of fibroin gene in a homologous cell-free system reveals an enhancing effect of the 5′-flanking sequence far upstream. Cell **27**:175–182.
41. **Vigneron, M., H. A. Barrera-Saldana, D. Baty, R. E. Everett, and P. Chambon.** 1984. Effect of the 21-bp repeat upstream element on *in vitro* transcription from the early and late SV40 promoter. EMBO J. **3**:2373–2382.
42. **Zarucki-Schulz, T., S. Y. Tsai, K. Itakura, X. Soberon, R. B. Wallace, M.-J. Tsai, S. L. C. Woo, and B. W. O'Malley.** 1982. Point mutagenesis of the ovalbumin gene promoter sequence and its effect on *in vitro* transcription. J. Biol. Chem. **257**:11070–11077.
43. **Zinn, K., D. Di Maio, and T. Maniatis.** 1983. Identification of two distinct regulatory regions adjacent to the human β-interferon gene. Cell **34**:865–879.