MOLECULAR AND CELLULAR BIOLOGY, Dec. 1986, p. 4259–4267
0270-7306/86/124259-09$02.00/0

Vol. 6, No. 12

Identification of Two Factors Required for Transcription of the Ovalbumin Gene

IKUKO SAGAMI, SOPHIA Y. TSAI, HENG WANG, MING-JER TSAI, AND BERT W. O'MALLEY*

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

Received 13 June 1986/Accepted 29 August 1986

Two transcription factors, COUP and S300-II, were isolated and partially purified from HeLa cell nuclear extracts. Both factors are required for the efficient transcription in vitro of the ovalbumin gene but not the simian virus 40 early genes. COUP factor binds to the chicken ovalbumin upstream promoter (COUP) sequence which lies between −70 to −90 base pairs upstream from the cap site. A series of competition experiments with a band-shifting assay was carried out to determine the relative affinity of COUP box transcription factor for various promoters. We found that a promoter DNA fragment isolated from the ovalbumin gene competes better than those isolated from the ovomucoid, Y, and α-actin genes. In contrast, the simian virus 40 early genes, the β-globin gene, and the adenosine deaminase gene promoters do not compete well in this assay. The molecular weight of the COUP factor was estimated by S-300 column chromatography, glycerol gradient centrifugation to be 90,000. However, two bands were observed in sodium dodecyl sulfate gel electrophoresis of cross-linked COUP factor to a ^{32}P-labeled oligonucleotide containing the COUP sequence. The protein moieties of the major and minor bands were estimated to be 85,000 to 90,000 and 40,000 to 45,000, respectively. The S300-II factor with an apparent molecular weight of 45,000 in an S-300 column is required for function in an in vitro reconstituted transcription system. In contrast to the COUP factor, the S300-II factor does not have apparent specificity for binding to the ovalbumin gene promoter. The S300-II factor may function by interacting with RNA polymerase or other DNA-binding transcription factors.

The interactions between specific DNA sequences, regulatory factors, and RNA polymerase II play an important role in modulating gene expression at the level of transcription. Many studies with DNA templates containing specific mutations have identified the *cis*-acting elements in DNA such as the conserved TATA box and the regions further upstream which are necessary for the efficient and accurate initiation of transcription. Recently, the GGGCGG sequence (GC box) located upstream of various genes was found to be important for their expression since deletion or mutation of this sequence resulted in a reduction or inactivation of their expression (14, 22, 23). Furthermore, Tjian and co-workers (7–10, 16, 19) have isolated a protein factor, SP1, which binds to the GC box and is essential for the transcription of this class of genes. Similarly, the CAAT box sequence (GG$^{C}_{T}$CAATCT consensus) which is *cis* linked and located around 80 base pairs upstream from the RNA initiation site of several genes (1, 11) has also been found to be important for its expression (4, 6, 18). In this case, a protein binding to that region has been identified in our laboratory and others, but its function has not been defined (12, 17, 27).

We have recently demonstrated that the distal promoter region of the chicken ovalbumin gene contains a duplicate GTCAAA box (GGTGTCAAAGGTCAAACT) which is essential for efficient and accurate transcription in both in vivo and in vitro systems (21, 27). We named this repeated sequence the chicken ovalbumin upstream promoter (COUP) element. In addition, in vitro transcription of the ovalbumin gene can be inhibited by the addition of an excess of competitor DNA fragments which contain the COUP box sequence. These experiments suggest the existence of a *trans*-acting factor which enhances transcription by binding to the COUP box sequence (27). Furthermore, footprinting with exonucleases or DNase I confirmed the presence of a COUP box binding protein in HeLa nuclear extracts (12, 27). The present study has been designed to characterize this COUP box factor and to elucidate its functional activity in transcription in vitro.

MATERIALS AND METHODS

Preparation of COUP transcription factors. Nuclear extracts from HeLa cells were prepared as described previously (27) and fractionated by using DEAE-Sephadex and phosphocellulose columns by the modified method previously defined for whole-cell extracts (32). The nuclear extract (50 ml) was loaded on a 180-ml DEAE-Sephadex (A25; Pharmacia) column preequilibrated with 20 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) (pH 7.9)–20% glycerol–0.1 mM EDTA–5 mM $MgCl_2$–2 mM dithiothreitol (buffer A) containing 50 mM $(NH_4)_2SO_4$. After the column was washed with the same buffer, proteins were eluted stepwise with 175 mM $(NH_4)_2SO_4$ and 500 mM $(NH_4)_2SO_4$ in buffer A. They were designated as DE50, DE175, and DE500, respectively. DE50 was dialyzed against 100 mM NaCl in buffer B (20 mM HEPES [pH 7.9], 20% glycerol, 0.2 mM EDTA, 2 mM dithiothreitol) and then loaded onto a 50-ml phosphocellulose column (P11; Whatman, Inc). The column was washed with 100 mM NaCl in buffer B, and proteins were eluted stepwise with 350, 500, and 1,000 mM NaCl in buffer B. Each fraction from the phosphocellulose column was designated as P100, P350, P600, and P1000, respectively. P600 was precipitated by the addition of $(NH_4)_2SO_4$ (60% saturation). This precipitate was dissolved in 1.0 ml of buffer A and dialyzed against buffer A containing 100 mM KCl to use for band-shifting, footprinting, and reconstituted transcriptional assays. The P600 (1.0 ml) was further chromatographed on a 72-ml Sephacryl S-300 column (Pharmacia) previously equilibrated with buffer A containing 100 mM KCl. The fractions having

* Corresponding author.

binding activity (S300-I or COUP box factor) and transcriptional activity (S300-II) were pooled. The S300-I fraction was concentrated to 0.5 ml by using Sephadex G-200 and then fractionated on a 33-ml S-300 column. The fractions containing binding activity were pooled and then loaded on a heparin-Sepharose CL6B (Pharmacia) column (1.0 ml) previously equilibrated with buffer A containing 100 mM KCl. After the column was washed with the same buffer, elution was carried out with a linear gradient of 50 to 400 mM KCl in buffer A. Fractions having the binding activity were pooled and dialyzed against buffer A containing 100 mM KCl. The S300-II fraction was also rechromatographed on a 33-ml S-300 column, and each fraction was assayed for the transcriptional activity in the presence of DE175, DE500, and S300-I fractions. Fractions having transcriptional activity (S300-II) were pooled and stored at −70°C.

Band-shifting assay. Deletion mutants (3′) of the ovalbumin gene constructed in our laboratory (L. Lopez, unpublished data) were digested with *Cla*I and end labeled with [α-^{32}P]dCTP (3,000 Ci/mmol; ICN Pharmaceuticals, Inc.) at the 3′ end and then digested with *Alu*I. Each ^{32}P-labeled DNA fragment (0.3 to 1 ng, 10^4 cpm) was incubated with a protein fraction and the appropriate amount of unlabeled *Hin*fI fragments of pBR322 at room temperature for 10 min in a reaction mixture containing 10 mM HEPES (pH 7.9), 100 mM KCl, 1 mM dithiothreitol, 0.05 mM EDTA, 2.5 mM $MgCl_2$, 6% glycerol, and 2% Ficoll (Pharmacia Fine Chemicals). The resulting complexes were then separated from protein-free DNA on a 5% polyacrylamide gel. The retardation of the mobility of DNA fragments was examined by exposure of the gel to XPR-1 X-ray film (Eastman Kodak Co.) with lighting plus intensifying screens.

Analysis of transcriptional activity. An in vitro reconstituted system and an RNase A mapping technique were used to analyze the accurate transcription of the ovalbumin gene as described previously (27). The reconstituted transcription system contains DE175, DE500, and P600 or each S300-I fraction as sources of transcriptional factors in 50 μl of reaction mixture containing 1.25 μg of the pSVOG DNA, 500 μM each of ATP, CTP, UTP, and GTP, 12 mM HEPES (pH 7.9), 3 mM $MgCl_2$, 60 mM KCl, 1.5 mM dithiothreitol, 0.06 mM EDTA, 2 mM spermidine, and 12% glycerol. The detailed construction of pSVOG has been described previously (21). The DNA template contains ovalbumin gene sequences from position −221 to position +41. The reaction mixture was incubated at 30°C for 45 min, and RNA synthesis was stopped by the addition of sodium dodecyl sulfate (SDS) to 0.05% and 250 μg of proteinase K. The RNA product was isolated and analyzed by RNase A mapping as described previously (27).

DNase I footprinting. DNase I footprinting analysis of each S-300 fraction was carried out as described previously (27). For DNase I footprinting of the coding strand, pSVOG (−161) was digested with *Cla*I, and the 3′ end was labeled with [^{32}P]dCTP and then redigested with *Pvu*II. This fragment contains sequences from −161 to +120. The ^{32}P-labeled *Cla*I-*Pvu*II fragment (1.4 ng) was incubated with 400 ng of pBR322 fragments and each protein fraction in a 20-μl reaction mixture containing 18 mM HEPES (pH 7.9), 90 mM KCl, 4.5 mM $MgCl_2$, 0.2 mM EDTA, 1.8 mM dithiothreitol, and 18% glycerol at room temperature for 5 min. A 0.14-μg amount of pancreatic DNase I (DPFF; Worthington Diagnostics) was added, and after incubation 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. The sample was prepared as described previously (27) and analyzed on 8% polyacrylamide–7 M urea sequencing gels.

UV cross-linking of the binding protein to an oligonucleotide. A double-strand oligonucleotide containing the COUP box (TCTATGGTGTCAAAGGTCAAACTTCTGA) with a 4-nucleotide overhang at each end was filled in with 100 μCi each of [^{32}P]dCTP, [^{32}P]dTTP, [^{32}P]dGTP, and [^{32}P]dATP by using the Klenow fragment to a specific activity of 5×10^8 cpm/μg. The S300-I fraction (0.17, 0.33, or 0.66 μg) was incubated with 0.2 ng of ^{32}P-labeled oligonucleotide and 500 ng of competitor *Hin*fI-pBR322 fragments at room temperature for 10 min. For cross-linking, samples were irradiated with UV light (115-W Mineralight lamp UVGL-25; Ultra-Violet Products, Inc.) for 20 min at a distance of 5 cm. The resulting complexes were separated from the unbound DNA on a 5% polyacrylamide gel and identified by exposure of the gel to X-ray film. The gel slices containing the protein-DNA complex were layered onto a 10% polyacrylamide-SDS protein gel with a stacking gel and run with protein markers. After electrophoresis, the gel was stained with Coomassie brilliant blue R-250 to visualize the protein markers. The gel was then dried and exposed to X-ray film at −70°C with intensifying screens.

Glycerol gradient ultracentrifugation. A 200-μl portion containing the S300-I fraction (4.3 μg) and SP6 polymerase (15 U) as an internal control was layered onto a 7 to 23% glycerol gradient (4.4 ml) containing 20 mM HEPES (pH 7.9), 5 mM $MgCl_2$, 0.1 mM EDTA, 0.1 M KCl, and 2 mM dithiothreitol. Centrifugation was performed for 16 h at 45,000 rpm in an SW50.1 rotor (Beckman Instruments, Inc.). Fractions of 6 drops (about 180 μl) were collected and analyzed for band-shifting activity and SP6 polymerase activity. To make a standard curve, a mixture of 10 μg each of sheep immunoglobulin G, bovine serum albumin, and ovalbumin was sedimented under similar conditions as described above except that the buffer did not contain dithiothreitol. The sedimentation rates of the standards were analyzed by running a sample of each fraction on an SDS gel followed by staining with silver reagent.

RESULTS

Detection of a COUP box transcription factor by the band-shifting assay. To facilitate the identification of the COUP box transcription factor throughout its purification, we used the band-shifting assay developed by Garner and Revzin (15) and Fried and Crothers (13). Initially, we fractionated HeLa nuclear extracts into three fractions on DEAE-Sephadex (DE50, DE175, and DE500) by the method of Tsai et al. (32). DE50 was then fractionated on a phosphocellulose column by elution with 100 mM (P100), 350 mM (P350), 600 mM (P600), and 1,000 mM (P1000) NaCl. The reconstituted transcription indicated that at least three fractions, DE175, DE500, and P600, were required for efficient in vitro transcription of the ovalbumin gene. DE175 contains the majority of the RNA polymerase II activity, while P600 was absolutely required for accurate initiation. When each of these six fractions was analyzed for its ability to bind to DNA by using DNase I footprinting assays, we found that most of the binding activity was contained in the P600 fraction (data not shown).

To utilize the band-shifting method to monitor the COUP binding factor, we incubated ^{32}P-labeled DNA fragments with the P600 fraction to form protein-DNA complexes. Incubations were carried out in the presence of increasing amounts of competitor DNA (pBR322) to reduce nonspecific

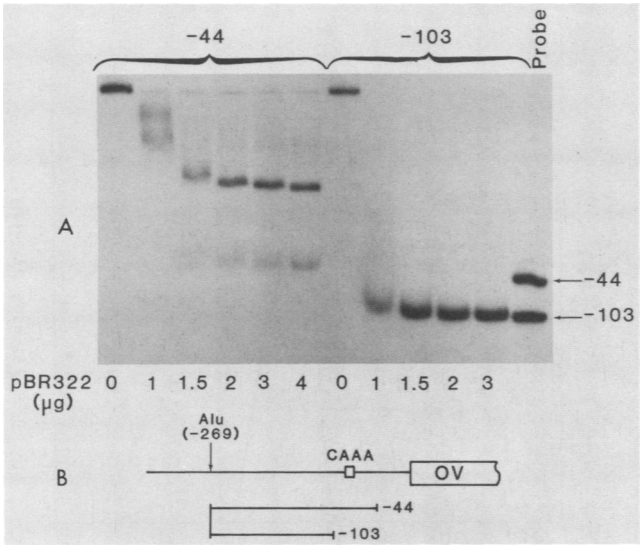


FIG. 1. (A) Binding of COUP box transcription factor to the ovalbumin promoter DNA in the presence of nonspecific DNA. (B) The end-labeled fragments contain sequences from −269 to −44 or −269 to −103 of the ovalbumin promoter. Each labeled DNA fragment (0.3 to 1 ng, 10^4 cpm) was incubated with a constant amount of P600 (2 μg) and increasing amounts of unlabeled competitor pBR322 under the conditions described in the text. Reactions were analyzed for binding activity on a low-ionic-strength gel.

protein-DNA interactions. The protein-DNA complex was then separated from protein-free DNA by virtue of the retarded mobility of the complex during electrophoresis in a native 5% polyacrylamide gel.

Two ^{32}P-labeled ovalbumin DNA fragments were used in these assays; one of the DNA fragments (−269 to −44) contained the COUP sequence, and the other fragment (−269 to −103) lacked this sequence (the duplicated GTCAAA sequence is located between −83 and −71). In the absence of competitor DNA, both DNA fragments bound to numerous proteins present in the P600 fraction, and these complexes failed to enter the gel (Fig. 1A). In the presence of competitor DNA, binding of the P600 fraction to the DNA fragment lacking the COUP sequence (−103) was not detectable, suggesting that the interaction of protein with this fragment was nonspecific. In contrast, the migration of the fragment containing the COUP sequence (−44) was retarded in the presence of competitor DNA. Since the shifted band was still detectable even at a nonspecific-to-specific DNA ratio of more than 4000:1, it indicates the significantly higher affinity of the P600 fraction factor for the COUP region of the ovalbumin gene promoter sequence. The factor(s) present in the P600 fraction was indeed a protein(s) since proteinase K or 100°C treatment prevented the formation of a shifted band (data not shown).

To identify the binding domain of the COUP transcription factor present in the protein-DNA complex, we carried out direct DNase I footprinting analysis of the complex. After incubation of the transcription factor with the end-labeled DNA (−44) probe of the ovalbumin promoter, DNase I was added to digest the unprotected DNA region, and the products were separated by native polyacrylamide gel electrophoresis. The shifted protein-DNA band was extracted from the gel and analyzed in a standard sequencing gel (Fig. 2). A region between −65 and −90 on the noncoding strand was protected from DNase I digestion. It was clear that the shifted band contained a factor which bound to the COUP box region of the ovalbumin gene promoter and that the band-shifting method is a valid assay system for detecting the COUP box binder.

Isolation of COUP transcription factor from HeLa nuclear extracts. For further purification of the COUP transcription factor, the P600 fraction was applied to a Sephacryl S-300 gel filtration column. Samples from the fractions of the S-300 column were incubated with ^{32}P-end-labeled DNA fragment containing a 3′ end at −44. In the presence of a constant amount of the labeled DNA fragment and pBR322 competitor, these assays identified a specific DNA-binding macromolecule which eluted between fractions 41 and 50 (Fig. 3A). As a control, when the −103 fragment was used in the

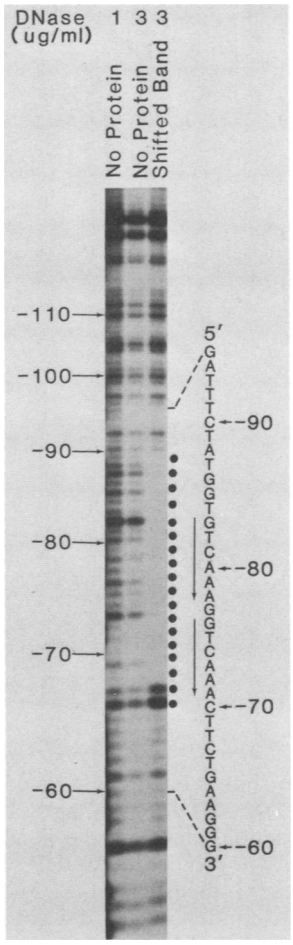
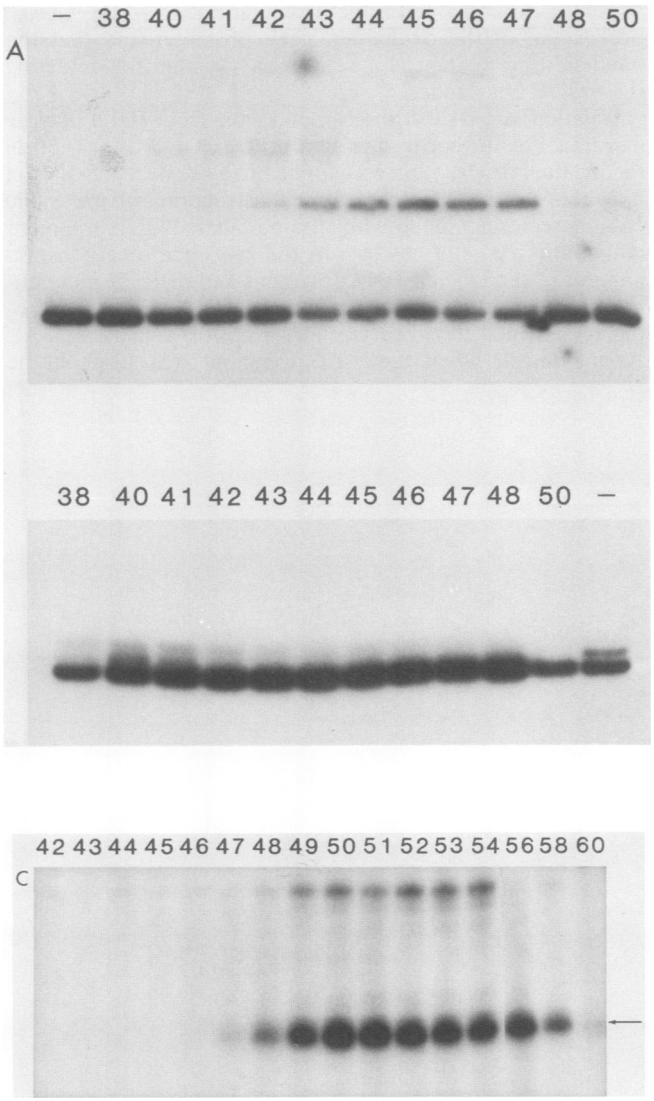


FIG. 2. DNase I footprinting of the COUP box transcription factor-DNA complex. A 12-μg amount of the P600 fraction was incubated with 3 ng (1.2×10^5 cpm) of end-labeled DNA fragment (−44) in the presence of poly(dI-dC) (20 μg) under the same conditions as described in the text. At the end of the incubation period, DNase I was added to a final concentration of 3 μg/ml and incubated at room temperature for 1 min. The reaction was terminated by the addition of EDTA to 10 mM. As a control, end-labeled DNA was digested in the absence of protein under identical conditions. Samples were loaded on a 5% polyacrylamide gel and protein-DNA complexes identified were excised from the gel. The DNAs were eluted by incubating gel slices in 1 ml of 0.5 M ammonium acetate containing 0.1% SDS, 0.1 mM EDTA, and 10 μg of proteinase K at 37°C overnight with shaking. The eluents were extracted with phenol-chloroform-isoamyl alcohol (25:24:1 [vol/vol]) and applied on an Elutip-d column, precipitated with ethanol, and then analyzed on an 8% sequencing gel.



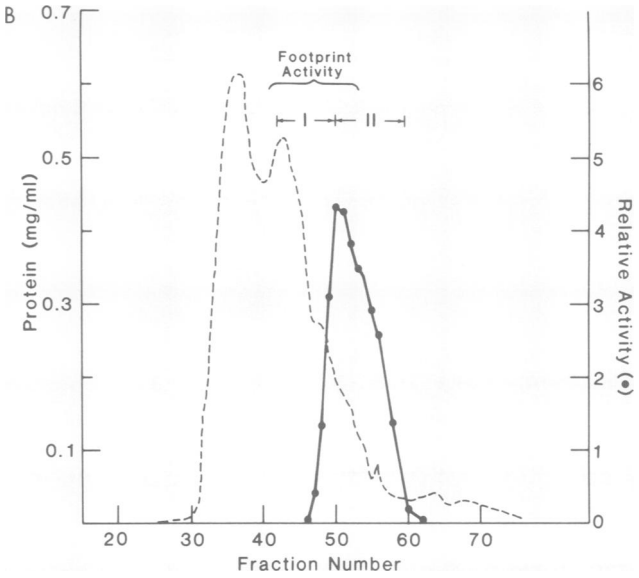


FIG. 3. Analysis of binding and transcriptional activities of the COUP box transcription factors in the S-300 fractions. (A) Assay for the presence of the COUP box transcription factor in the first S-300 column fractions. A 1-μl volume of each of the indicated fractions containing 0.2 to 0.6 μg of protein was incubated with 0.6 ng of ^{32}P-end-labeled fragment −269 to −44 (top) or −269 to −103 (bottom) of the ovalbumin promoter, and 1 μg of *Hin*fI-digested pBR322 DNA fragments. (B) Sephacryl S-300 gel filtration of P600. The P600 fraction was loaded on a Sephacryl S-300 column (72 ml), and each 20 drops was collected. Relative activity of transcription (●—●) was estimated by densitometric scanning of the lower exposure autoradiogram shown in panel C. Protein (– – –) was estimated by Bradford assays. DNase I footprinting analysis was also carried out, using 18 μl of each fraction. (C) Analysis of transcriptional activities of the S-300 fraction. An in vitro reconstituted transcriptional system was carried out with 2.5 μl of DE175, 1.5 μl of DE500, and 15 μl of each S-300 fraction as sources of transcription factors and the ovalbumin gene (pSVOG) as a template under the conditions described in the text. The RNA product was analyzed by RNase A mapping as described in the text. The arrow shows the position of the accurately initiated transcript (120 nucleotides).

binding assay, no shifted band could be detected (Fig. 3A). Similarly, when DNase I footprinting assays were carried out, the same fractions that had the band-shifting activity also had the ability to protect the COUP sequences (−65 to −92) from DNase I digestion (Fig. 3B). However, the reconstituted transcription assays did not correlate exactly with the band-shifting and footprint activity in that a peak of transcriptional activity was detected at around fraction 51 (Fig. 3B and C). These results suggested that the P600 fraction might contain two separate activities, a COUP sequence specific binding activity and an additional factor necessary for transcription.

To verify the presence of these two distinct factors, fractions with the COUP binding activity (S300-I) and fractions with the transcriptional activity (S300-II) were rechromatographed on another S-300 column (Fig. 4). By using band-shifting and DNase I footprinting analysis, we detected the COUP binding activity only in those fractions obtained by rechromatography of S300-I (Fig. 4A and B). Each fraction alone did not have any transcriptional activity in the reconstituted system. However, when transcription assays were carried out in the presence of fraction 57 obtained from rechromatography of the S300-II fraction (Fig. 4C), we found a peak of activity around fractions 41 and 42. To avoid the contamination of COUP binding factor in the S300-II fractions, in this particular experiment we used fraction 57, which was located on the distal side of the peak fractions having the COUP binding activity. The transcriptional activity correlated well with the binding activity as measured by DNase I footprinting and band-shifting techniques. Thus, both S300-I and S300-II are required for efficient transcription of the ovalbumin gene.

Similarly, S300-II fractions alone supported only a low level of transcriptional activity (Fig. 4C), but the addition of fraction 41 of S300-I, which contained the COUP binding factor, resulted in a significant increase in the transcriptional activity of the S300-II fractions (Fig. 4C and D). These results indicated that at least two distinct transcription factors were contained in the P600 fraction which could be separated by chromatography on a gel filtration column. One is the COUP box transcription factor which has a molecular weight of about 90,000 in its native form (estimated by S-300 gel filtration and assuming the factor has a globular structure). By using the same assumptions, the other factor,

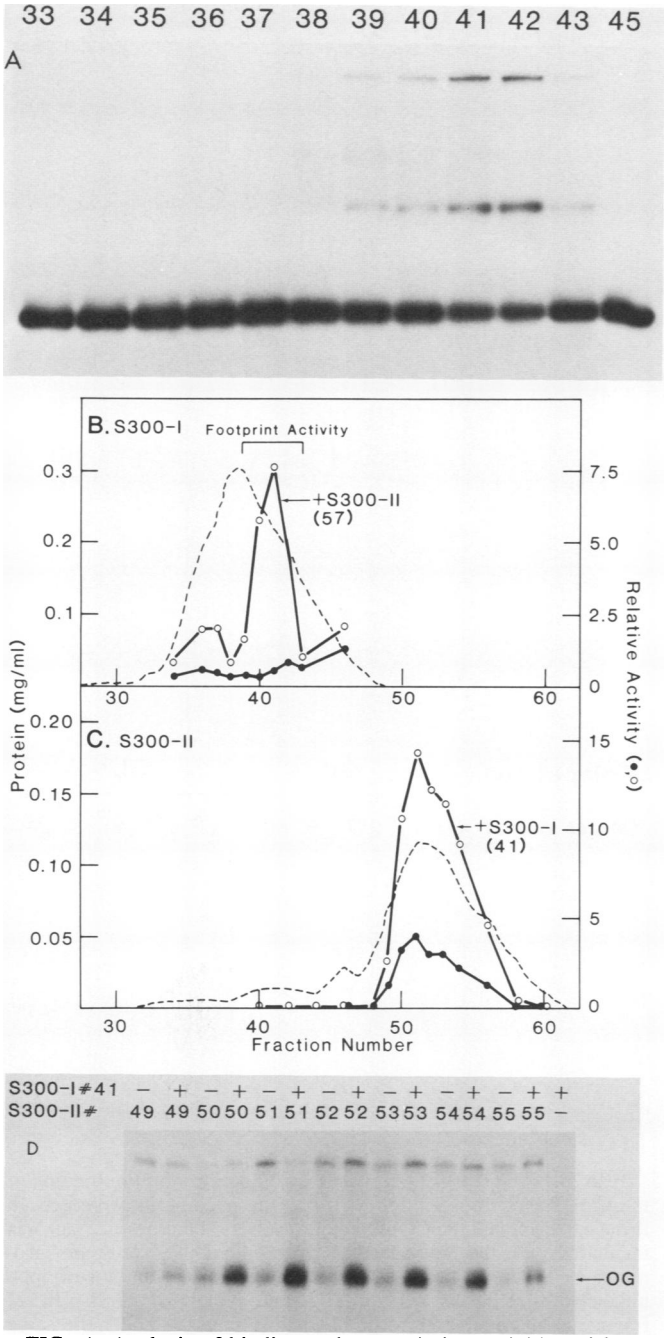


FIG. 4. Analysis of binding and transcription activities of fractions rechromatographed in a Sephacryl S-300 column. (A) Assay for the presence of the S300-I transcription factor in the rechromatographed S:300 fractions. Each of the indicated fractions containing 0.15 to 0.45 μg of protein (1.5 μl) was incubated with 0.6 ng of a ^{32}P-end-labeled DNA fragment (−269 to −44) in the presence of 500 ng of *Hin*fI-pBR322 fragments and analyzed as described in the legend to Fig. 1. (B) Rechromatography of fraction I on a second S-300 column. A 7.5-μl volume of each rechromatographed S300-I fraction was analyzed for transcription of pSVOG in the reconstituted system containing 2.5 μl of DE175 and 1.5 μl of DE500 in the absence (●—●) or in the presence (○—○) of rechromatographed S300-II (fraction 57, 15 μl) obtained from the rechromatography of the S300-II fraction. DNase I footprinting analysis also was carried out with 18 μl of each fraction. Protein (----) was estimated by Bradford assays. (C) Rechromatography of the S300-II fraction on a second S-300 column. A 15-μl volume of each fraction was analyzed for its transcriptional activity in the reconstituted system in the absence (●—●) or in the presence (○—○) of rechromatographed S300-I (fraction 41, 7.5 μl) under the same conditions as described in the text. (D) In vitro transcription of the ovalbumin gene with S300-II fractions in the absence (−) or presence (+) of the COUP box transcription factor (S300-I, fraction 41). The reactions were carried out as described in the text.

S300-II, has a lower molecular weight (about 45,000). Each factor appears to be necessary for efficient transcription of the ovalbumin gene. We should emphasize that in the absence of the S300-I fraction some basal level of transcriptional activity was still observed.

The S300-I fractions having binding activity to the COUP sequence were pooled and purified further through a heparin-Sepharose column. The bound protein was eluted with a linear gradient of KCl from 0.05 to 0.4 M in buffer A. The COUP binding activity was eluted at a salt concentration around 0.3 M KCl (data not shown). The COUP and S300-II factors isolated this way were substantially purified but still contained several bands in an SDS silver-stained gel. The minimal purification of the COUP box transcription factor was 300-fold from the HeLa nuclear extract. The level of purification of the COUP box factor was estimated by quantitation of the shifted band by using each protein fraction. The estimated purity of the COUP box binding protein at this stage was 1 to 10%, assuming that the COUP box transcription factor has a molecular weight of 90,000 and that one COUP box binding protein binds to one molecule of DNA probe. We should emphasize that this is a minimal estimation since some of the COUP binders certainly bind to the competitor DNA and are not detected.

Characterization of the COUP binding protein. The molecular weight of the COUP box transcription factor was estimated to be about 90,000 by a gel filtration column. In addition, we carried out glycerol gradient ultracentrifugation to estimate the molecular weight of the COUP box factor (Fig. 5A). The activity of COUP box factor was localized in fractions by the band-shifting assay. As an internal standard, SP6 polymerase was cosedimented, and each fraction was analyzed for polymerase activity. The molecular weight estimated this way was ca. 85,000. Thus the "native" COUP box binder is more or less a globular structure. To identify the COUP box transcription factor on an SDS-polyacrylamide gel, we used a ^{32}P-labeled oligonucleotide (24-mer) containing the COUP box DNA sequence as a probe to label the COUP factor. After cross-linking the protein to the oligonucleotide with UV irradiation, the protein-DNA complexes were separated from protein-free oligonucleotides on a 5% polyacrylamide gel. The gel slices containing the protein-DNA complex were excised and analyzed on a 10% polyacrylamide SDS gel. Three bands were seen (Fig. 5B). The low band corresponds to free DNA, and the intensity of the upper two bands increased with increasing amounts of the S300-I fraction. The molecular weights of these bands were estimated to be about 60,000 to 105,000. Since the attached oligonucleotide has a molecular weight of about 12,000, the molecular weight of the major protein band was estimated to be 42,000 and that of the minor band was estimated to be 87,000 on this SDS protein gel. Similarly, in a control experiment, we cross-linked nonspecifically SP6 RNA polymerase (98,000) to the same oligonucleotide under the identical conditions except that no competitor *Hin*fI-pBR322 DNA fragments were added during the formation of the cross-linked complex. In this way, we found that the nonspecific complex of SP6 RNA polymerase-oligonucle-

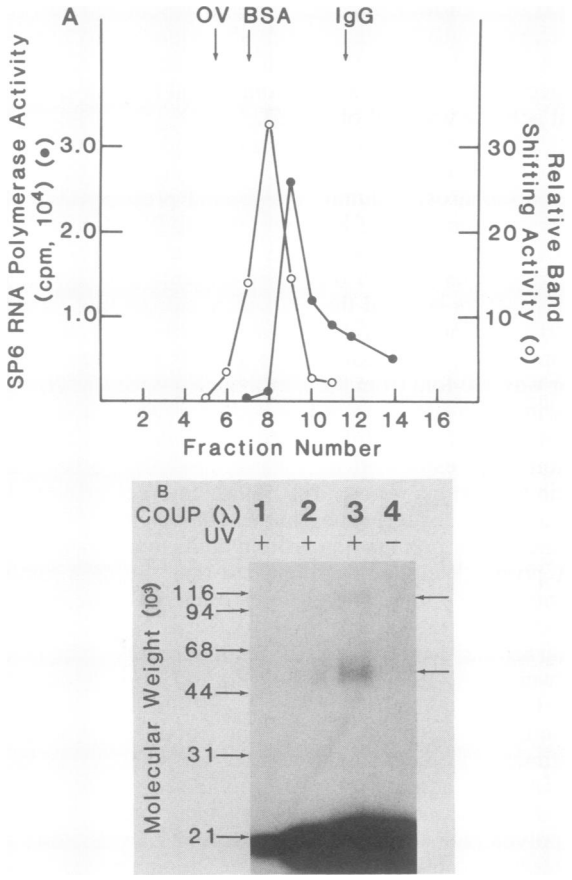


FIG. 5. (A) Glycerol gradient analysis of the COUP box transcription factor. COUP box transcription factor was sedimented as described in the text. A 5-μl volume from each fraction was assayed for COUP box binding activity as described in Fig. 1. SP6 polymerase (98,000 molecular weight) activity was measured by the procedure of the supplier. The molecular weight standards were detected by silver staining of the SDS-polyacrylamide gel: OV, Ovalbumin (44,000); BSA, bovine serum albumin (68,000); IgG, sheep immunoglobulin G (158,000). (B) Identification of the COUP box binding factor on an SDS-polyacrylamide gel by UV cross-linking. A cross-linked COUP box transcription factor-^{32}P-oligonucleotide complex was formed by using UV irradiation in the presence of 500 ng of *Hin*fI-pBR322 DNA fragments and separated from protein-free oligonucleotide in an acrylamide gel as described in the text. The gel slices containing the protein-DNA complexes were loaded onto a 10% polyacrylamide-SDS gel and analyzed. Lanes: 1, 0.17 μg of COUP box transcription factor cross-linked to 0.2 ng of ^{32}P-oligonucleotide by UV irradiation; 2, 0.33 μg of COUP box transcription factor cross-linked to 0.2 ng of ^{32}P-labeled oligonucleotide by UV irradiation. 3, 0.66 μg of COUP box transcription factor cross-linked to 0.2 ng of ^{32}P-oligonucleotide by UV irradiation; 4, 0.66 μg of COUP box transcription factor and 0.2 ng of ^{32}P-oligonucleotide. Arrows indicate protein-DNA complexes.

otide moved as a 115,000-molecular-weight complex (data not shown). These results indicate that the native S300-I factor may exist as a dimer of 40,000- to 45,000-molecular-weight subunits.

Specificity of the COUP box transcription factor. To determine the specificity of the COUP box transcription factor for various promoter sequences, a series of competition studies was performed with band-shifting assays by using unlabeled promoter DNA fragments of several genes as competitors (Fig. 6A to C). The binding of the COUP box transcription

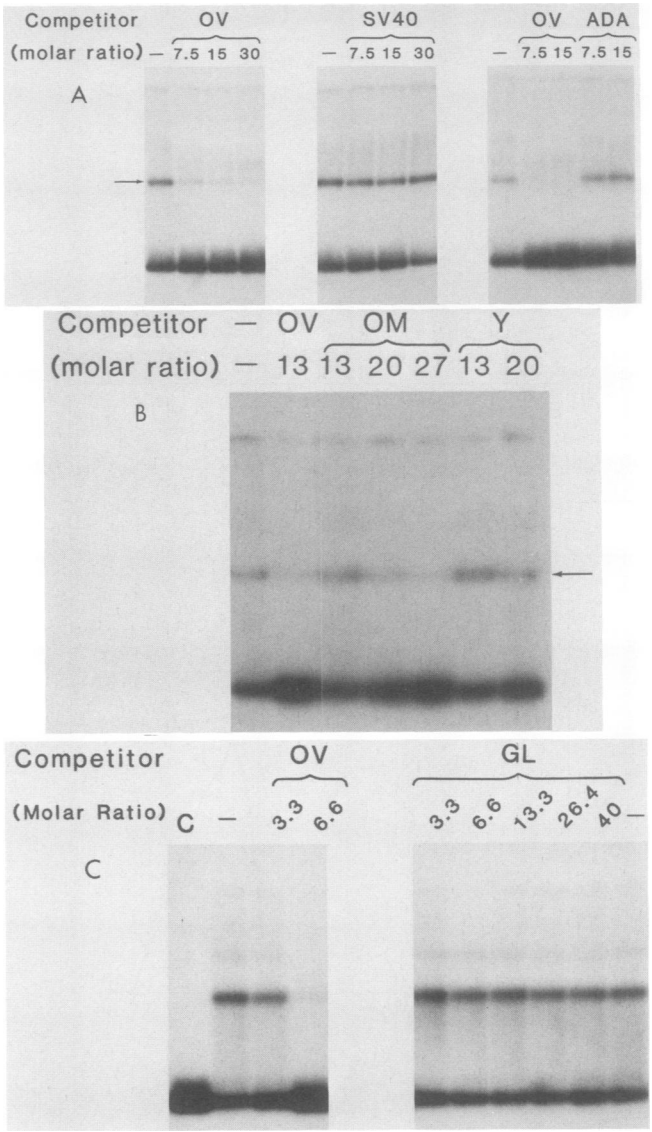


FIG. 6. (A and B) Effect of DNA competitors on the binding of the COUP box transcription factor to the ovalbumin promoter. In these experiments, the COUP box transcription factor (0.25 μg) was incubated with the labeled DNA fragment of the ovalbumin promoter (−269 to −44, 0.6 ng), with an increasing amount of each unlabeled specific competitor DNA and 500 ng of *Hin*fI-pBR322 fragments and analyzed in the band-shifting assay under the same conditions as described in the text. The DNA fragment of the ovalbumin gene (OV) which was used as a competitor contains upstream sequences from −753 to −56 was prepared as described previously (27). The fragment of the SV40 early genes contains from −340 to +150 (*Pvu*II-*Pvu*II 490-base-pair fragment), and the adenosine deaminase gene (ADA) contains from ca. −150 to +1,200 (*Eco*RI to *Eco*RI 1.35-kilobase fragment). The *Hin*dIII-*Hin*dIII fragment (−1,600 to +263, 1.9 kilobases) of the ovomucoid gene (OM) and the *Pst*I-*Hin*dIII fragment (−1,456 to +619, 2.1 kilobases) of the Y genes also were prepared as competitor DNAs. (C) Effect of the β-globin gene promoter on the binding of the COUP box transcription factor to the ovalbumin promoter. The competition experiments in the band-shifting assay were carried out under the same conditions as described for panels A and B. The DNA fragment of the β-globin gene (GL) used as a competitor is from −205 to +75 (*Nco*I-*Hin*fI, 280-base-pair fragment). Minus signs indicate no specific competion in the reaction mixture.

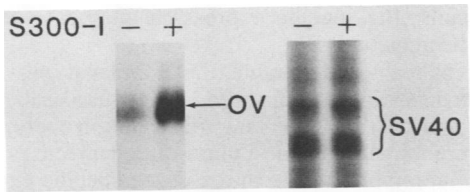


FIG. 7. COUP box transcription factor is required for the ovalbumin gene but not for the SV40 early gene transcription. The transcription of the ovalbumin and SV40 early genes was examined by using an in vitro reconstituted system with 2.5 μl of DE175, 1.5 μl of DE500, 5 μl of S300-II (fragment 52) in the absence (−) or in the presence (+) of the COUP box transcription factor (6 μl, fragment 42) as described in the text. A 1.25-μg amount of pSVOG was used as a template for the synthesis of ovalbumin and SV40 RNA. The detailed construction of pSVOG and the [^{32}P]RNA probe made from pSPOG and pSPSV were described in our previous paper (27). The arrow and the bracket indicate the position of the RNase A mapping products. The correctly initiated transcripts from the ovalbumin gene and the SV40 early genes will protect 120 nucleotides and 60 to 70 nucleotides of probe, respectively.

factor to the COUP box region of the ovalbumin gene promoter (−269 to −44) was totally competed with ovalbumin DNA fragment (−753 to −56) at a molar ratio of competitor-to-labeled fragment of 13 but not with the simian virus 40 (SV40) early gene promoters (−340 to +150) (Fig. 6A) at the same or higher molar ratios. Similarly, the promoter fragment of the adenosine deaminase gene (ADA) which lacks a GTCAAA sequence had no effect on the binding of the COUP box transcription factor to the ovalbumin promoter (Fig. 6A). In contrast, DNA fragments of the ovomucoid gene and the Y gene competed for binding (Fig. 6B), although at a lesser efficiency than that of the ovalbumin promoter. The ovomucoid gene promoter has a GTCCAACGAT sequence at the position of −56. The Y gene promoter has a TGGCAAAGAC sequence at the position of −81 and also has a TGCCAAACTG sequence at −77 in a reverse orientation. Similarly, the rat insulin genes I and II and the chicken α-actin gene promoter fragments also competed for the binding to COUP box transcription factor but at an even lower efficiency (data not shown). Sequence comparison of these genes shows that there are similar sequences of GNCAAAGA at the −50 to −90 region for binding to COUP box factor. These results suggest that the COUP box transcription factor is specific for a group of genes which contain the COUP box consensus within their promoter. In addition, these results indicate that different promoters may have different binding affinities for the COUP box transcription factor because of variations in their surrounding sequences.

Since the COUP box sequences have some sequence similarity to the CAAT consensus GGCCAATCT, we also carried out competition experiments with the chicken β-globin gene promoter as a competitor. The β-globin gene promoter contains a perfect CAAT consensus at the −77 to −72 position. The β-globin gene promoter, even at a molar ratio of 40:1, has no effect on the binding of COUP box transcription factor to its binding site (Fig. 6C). In contrast, the ovalbumin gene promoter completely abolished this binding at a sixfold molar ratio of competitor. This result indicated that the COUP box transcription factor is not the same factor that bound to the CAAT consensus of the β-globin gene.

Next, we carried out reconstituted transcription assays by using the ovalbumin gene and the SV40 early genes as templates. As expected, the presence of the COUP box transcription factor in a reconstituted system stimulated the transcription of the ovalbumin gene (Fig. 4C and 7). On the other hand, the transcription of the SV40 early genes which contains no GGNCAAAGAG consensus in its promoter was not affected by the presence of the S300-I fraction. These results once again suggest that the factor in the S300-I fraction may be required for the transcription of genes which contain the COUP box consensus in their promoter sequence but not for genes lacking it.

DISCUSSION

In the present paper, we describe the fractionation of HeLa cell nuclear extracts and the isolation of two transcription factors, COUP box factor (S300-I) and S300-II. DNase I footprinting analysis demonstrated that SI binds specifically to the promoter region of the ovalbumin gene (COUP). This binding site contains a direct repeat of the GTCAAA sequence.

Several lines of evidence suggested that the COUP box transcription factor is a necessary factor for efficient transcription of the ovalbumin gene. First, COUP binding activity was copurified with the transcriptional activity through a number of steps of purification. Both band-shifting and DNase I footprinting data indicated that the binding activities correlate well with the transcriptional activity in S-300 and heparin-Sepharose column chromatography. Second, our reconstituted transcription studies revealed that the isolated COUP box transcription factor is required for efficient transcription of the ovalbumin gene. In contrast, the transcription of the SV40 early genes lacking this sequence in its promoter does not require the COUP box transcription factor. In addition, other promoter DNA fragments containing the related sequence, such as ovalbumin, ovomucoid, Y, chicken α-actin, and rat insulin genes, can compete for the binding of the COUP box transcription factor, whereas promoter DNA fragments lacking the sequence, such as the SV40 early genes and the adenosine deaminase gene, cannot compete for the binding of this factor. Finally 5′ deletion mutants which have the COUP box transcription factor binding site deleted have also lost their transcriptional activity (27). Taken together, these results suggest that the binding of COUP box transcription factor to the ovalbumin gene upstream promoter sequence is a necessary prerequisite for efficient transcription.

In the present paper, we report the isolation of another transcription factor, S300-II. This S300-II factor is also essential for the transcription of the ovalbumin gene. The footprinting and band-shifting data indicated that S300-II alone does not bind to the COUP region of the ovalbumin gene promoter. The S300-II factor might function by interaction with RNA polymerase II or other transcription factors. In this aspect, it is interesting to note that by using an anti-RNA polymerase antibody affinity column we identified a transcription fraction (SE) which interacts with RNA polymerase (31). This protein fraction can be substituted by the P1000 or P600 fraction from which COUP box and S300-II factors were isolated. Thus, the S300-II factor may associate with RNA polymerase II. If this is the case, S300-II could be considered somewhat analogous to the σ-like factors in procaryotic RNA polymerase transcription systems. Further studies are needed to understand these molecular events.

Since HeLa cells contain the COUP box transcription

factor which can bind the function with chicken genes, it indicates the high degree of conservation of the protein factor and the COUP box sequence in eucaryotes. Recently, we also identified a similar factor in oviduct nuclear extracts which binds specifically to the COUP box of these same genes (unpublished data). This oviduct fraction containing the COUP box binding factor can substitute for the HeLa cell S300-I fraction in the in vitro transcription of the ovalbumin gene. While we were preparing the present paper, Jones et al. also reported a factor (CTF) in HeLa cells that binds to the CAAT box region of the herpesvirus thymidine kinase gene (20). Similarly, Graves et al. (17) and Cohen et al. (5) have identified a CAAT box binding factor in soluble rat liver extracts and murine erythroleukemia cell extracts. Since the COUP box has some sequence homology to the CAAT box consensus (GTGTCAAAGG versus GGCTCA ATCT), it was possible that the COUP transcription factor is the same as those reported. However, DNA fragments containing the CCAAT sequence from the chicken β-globin gene promoter cannot compete for the binding of S300-I to the COUP box sequence of the ovalbumin gene promoter. In addition, we also detected the binding activity to the β-globin gene in fractions of the first S-300 column which separated from the fractions containing the binding activity to the COUP box (data not shown). These results suggest that the COUP box binding protein is different from the CAAT box binding protein for the β-globin gene. Thus, it is likely that the COUP box transcription factor represents either another class of transcription factors or a distinct subclass of a larger family of CAAT box transcription factors which differ either by peptide sequence or posttranslational modification.

Several reports have suggested that common transcription factors are shared by several genes. These include the TATA box transcription factor, which has been shown to be required by TATA box-containing genes (28; R. G. Roeder, personal communication). The factor SP1 is required by genes containing the GGGCGG (GC box) sequence (7–10, 16, 19), the CAAT box transcription factor is required by genes containing CCAAT box (5, 17, 20), and the heat shock-specific transcription factor is required by heat shock genes (25, 26, 30, 33, 34). Most recently, the transcription factor MLTF was shown to bind to the upstream element in the adenovirus major late promoter as well as to several other gene promoters (3, 24, 28). We demonstrated that the COUP box transcription factor is required for a class of genes containing the COUP box-related sequences. Such factors are common to multiple genes but appear to be specific for different classes of genes. It is interesting that the herpesvirus thymidine kinase gene contains at least three classes of transcription regulators which use the TATA, CAAT, and GC boxes; all of these three elements are essential for transcriptional activity (17, 19). Thus, it is probable that all three *trans*-acting factors (i.e., SP1, TATA, and CAAT box transcription factors) contribute toward its expression in vivo.

How *trans*-acting factors regulate the tissue-specific expression of genes is unknown at present. The results of Brown (2) may shed some light on this important question. Oocyte 5S genes are expressed only in oocytes but not in somatic tissues. In contrast to the common expectation, this developmentally specific expression of oocyte 5S genes is controlled by a common transcription factor, TFIIIA, which exists in both cell types. Through a differential binding affinity of TFIIIA to somatic and oocyte genes and by variations in the concentration of TFIIIA in oocyte and somatic cells, the specific expression of oocyte genes appears to be achieved.

It is reasonable to assume that similar types of gene regulation may exist also in polymerase II-transcribed genes. If this is the case, COUP box transcription factors which have differential affinity for a class of eucaryotic promoters may play an important role in the tissue-specific expression of this same gene set. In this regard, it is interesting that a factor has been detected which binds specifically to the octanucleotide sequence ATTTGCAT of the immunoglobulin genes (29). This octanucleotide sequence has been determined to be important for the tissue-specific expression of immunoglobulin genes. However, this factor can be detected in both B and HeLa cell extracts. Although the exact role of this binding factor in the tissue-specific expression of immunoglobulin genes is yet to be defined, it is possible that tissue-specific expression may not require that such binding factors be expressed only in the homologous cell. As mentioned earlier, six "common" transcription factor (TATA box, CAAT box, GC box, heat shock MLTF, and COUP transcription factors) have already been identified cell extracts. It is possible that there are many more common transcription factors existing in eucaryotic cells. Since different genes use different sets of common regulatory elements,it is conceivable that through the differential utilization of a defined set of common regulatory elements and by variation in the level of their *trans*-acting factors, eucaryotes promote both the rate and selectivity of gene expression.

ACKNOWLEDGMENTS

We thank Peter Hine and Geoffrey Long for their excellent technical assistance.

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. **Brown, D. D.** 1984. The role of stable complexes that repress and activate eucaryotic genes. Cell **37**:359–365.
3. **Carthew, R. W., L. A. Chodosh, and P. A. Sharp.** 1985. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell **43**:439–448.
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. **Cohen, R. B., M. Sheffery, and C. G. Kim.** 1986. Partial purification of a nuclear protein that binds to the CCAAT box of the mouse α_1-globin gene. Mol. Cell. Biol. **6**:821–832.
6. **Dierks, P., A. van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissman.** 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. **Dynan, W. S., J. D. Saffer, W. S. Lee, and R. Tjian.** 1985. Transcription factor SP1 recognizes promoter sequences from the monkey genome that are similar to the simian virus 40 promoter. Proc. Natl. Acad. Sci. USA **82**:4915–4919.
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. **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.
13. **Fried, M., and D. M. Crothers.** 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9**:6505–6525.
14. **Fromm, M., and P. Berg.** 1982. Deletion mapping of DNA required for SV40 early promoter function *in vivo*. J. Mol. Appl. Genet. **1**:457–481.
15. **Garner, M. M., and A. Revzin.** 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions. Nucleic Acids Res. **9**:3047–3060.
16. **Gidoni, D., W. S. Dynan, and R. Tjian.** 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. Nature (London) **312**:409–413.
17. **Graves, B. J., P. F. Johnson, and S. L. McKnight.** 1986. Homologous recognition of a promoter common to the MSV LTR and the HSV tk gene. Cell **44**:565–576.
18. **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.
19. **Jones, K. A., and R. Tjian.** 1985. Sp1 binds to promoter sequences and activates HSV "immediate-early" gene transcription *in vitro*. Nature (London) **317**:179–182.
20. **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.
21. **Knoll, B. J., T. Zarucki-Schultz, 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.
22. **McKnight, S. L., and R. C. Kingsbury.** 1982. Transcriptional control signals of a eucaryotic protein coding gene. Science **217**:316–324.
23. **McKnight, S. L., R. C. Kingsbury, A. Spence, and M. Smith.** 1984. The distal transcription signals of the herpes virus tk gene share a common hexanucleotide control sequence. Cell **37**:253–262.
24. **Miyamoto, N. G., V. Moncollin, J. M. Egly, and P. Chambon.** 1985. Specific interaction between a transcription factor and the upstream element of the adenovirus-2 major late promoter. EMBO J. **4**:3563–3570.
25. **Parker, C. S., and J. Topol.** 1984. A *Drosophila* RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. Cell **36**:357–369.
26. **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.
27. **Pastorcic, M., H. Wang, A. Elbrecht, S. Y. Tsai, M.-J. Tsai, and B. W. O'Malley.** 1986. Control of transcription initiation in vitro requires binding of a transcription factor to the distal promoter of the ovalbumin gene. Mol. Cell. Biol. **6**:2784–2791.
28. **Sawadogo, M., and R. G. Roeder.** 1986. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43**:165–175.
29. **Singh, H., R. Sen, D. Baltimore, and P. A. Sharp.** 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature (London) **319**:154–158.
30. **Topol, J., D. M. Ruden, and C. S. Parker.** 1985. Sequences required for *in vitro* transcriptional activation of a *Drosophila* hsp 70 gene. Cell **42**:527–537.
31. **Tsai, S. Y., P. Dicker, P. Fang, M.-J. Tsai, and B. W. O'Malley.** 1984. Generation of monoclonal antibodies to RNA polymerase II for the identification of transcription factors. J. Biol. Chem. **259**:11587–11593.
32. **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.
33. **Wu, C.** 1984. Two protein-binding sites in chromatin implicated in the activation of heat-shock genes. Nature (London) **309**:229–234.
34. **Wu, C.** 1984. Activating protein factor binds *in vitro* to upstream control sequences in heat-shock gene chromatin. Nature (London) **311**:81–84.