MOLECULAR AND CELLULAR BIOLOGY, Jan. 1988, p. 105–113
0270-7306/88/010105-09$02.00/0

Vol. 8, No. 1

DNA Affinity Labeling of Adenovirus Type 2 Upstream Promoter Sequence-Binding Factors Identifies Two Distinct Proteins

BRIAN SAFER,* ROGER B. COHEN, SUSAN GARFINKEL, AND JOHN A. THOMPSON

Section on RNA and Protein Biosynthesis, Laboratory of Molecular Hematology, National Heart, Lung, and Blood Institute, Bethesda, Maryland 20892

Received 14 July 1987/Accepted 14 October 1987

A rapid affinity labeling procedure with enhanced specificity was developed to identify DNA-binding proteins. ^{32}P was first introduced at unique phosphodiester bonds within the DNA recognition sequence. UV light-dependent cross-linking of pyrimidines to amino acid residues in direct contact at the binding site, followed by micrococcal nuclease digestion, resulted in the transfer of ^{32}P to only those specific protein(s) which recognized the binding sequence. This method was applied to the detection and characterization of proteins that bound to the upstream promoter sequence (−50 to −66) of the human adenovirus type 2 major late promoter. We detected two distinct proteins with molecular weights of 45,000 and 116,000 that interacted with this promoter element. The two proteins differed significantly in their chromatographic and cross-linking behaviors.

The accurate and regulated expression of genes transcribed by RNA polymerase II requires the interaction of specific DNA-binding proteins with *cis*-acting promoter elements (2, 10, 14, 29, 36). Several well-characterized methods have been developed which readily identify the locations, boundaries, and sequences of these regulatory regions (11–13, 41). A more difficult problem, however, has been to identify the DNA-binding proteins that specifically recognize these sites and to characterize the mechanisms by which they modulate the activity of the transcription unit. At present, identification of these *trans*-acting factors requires extensive purification to near or apparent homogeneity. In addition to recognizing specific DNA sequences, the purified factor must also affect the transcriptional activity of the gene with which it interacts.

We developed three interdependent strategies to rapidly identify, purify, and characterize sequence-specific DNA-binding proteins which regulate gene expression. First, these factors were extensively purified by ion-exchange and DNA Mono Q (Pharmacia) affinity chromatography with multicopy promoter elements (R. B. Cohen, S. Garfinkel, J. A. Thompson, and B. Safer, manuscript in preparation). Second, function was monitored in a novel assay system generated by the addition of specific DNA fragments to HeLa whole-cell extract (WCE). This produced a functional sequestration of the cognate transcription factor in that it was bound to the DNA fragment. The activity of the factor was then assayed by restoring function to the depleted extract with the purified protein (S. Garfinkel, R. B. Cohen, J. A. Thompson, and B. Safer, manuscript in preparation). Third, the DNA-binding proteins that recognized these sequences were identified by a novel modification of the UV-dependent cross-linking procedure (17, 25, 27, 31) which involved the transfer of ^{32}P from specific phosphodiester bonds at the binding site to the polypeptide chain. When analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE), the M_r of the binding protein could be determined directly. The third procedure is the subject of this report. We applied these procedures toward the purification of factors that interact with the upstream promoter sequence (UPS) of the adenovirus type 2 (Ad2) major late promoter (MLP). Results of deletion and mutation analyses have shown that efficient transcription of the Ad2 MLP is dependent on an imperfect palindrome and flanking sequences extending from positions −50 to −66 (15, 18, 20). Subsequently, a single 45- to 55-kilodalton (kDa) protein that bound to this region has been extensively purified in several laboratories (6, 30, 35). The data presented here, however, demonstrate that two distinct polypeptides bind to this positive regulatory sequence.

* Corresponding author.

MATERIALS AND METHODS

Mono S fractionation of HeLa WCE. HeLa WCE (26) or K562 nuclear extract (9) was dialyzed against buffer A [40 mM $(NH_4)_2SO_4$, 6 mM $MgCl_2$, 20 mM Tris hydrochloride (pH 7.9), and 1 mM dithiothreitol], clarified by centrifugation at 100,000 × *g* for 20 min, and chromatographed on Mono S cation-exchange resin on a high-performance liquid chromatographic system (Pharmacia Fine Chemicals, Piscataway, N.J.). Following a wash with buffer A, protein was eluted with a 45-ml KCl gradient from 0 to 0.6 M KCl in buffer A. Fractions of 2 ml were concentrated to 55 μl with a concentration unit (Centricon 30; Amicon Corp., Lexington, Mass.). Fractions were assayed immediately or were stored in the vapor phase of liquid N_2. Activity was stable for periods of up to 6 months.

^{32}P-labeled UPS DNA probe construction. We constructed DNA probes containing ^{32}P at specific phosphodiester bonds within the Ad2 MLP UPS by using three related procedures. The first method was applicable to DNA fragments that contained restriction endonuclease recognition sequences within the binding site. A single internal phosphodiester bond on each strand was radiolabeled by this procedure (Fig. 1).

In this procedure *Ava*I digestion of the multicopy plasmid pMC347×6 was used (Garfinkel et al., in preparation) to generate a 347-base-pair (bp) Ad2 MLP promoter fragment that extended from positions −138 to +198 with *Ava*I linkers on both ends. Following purification by PICS chromatography (43), the 347-bp fragment was then subjected to restriction endonuclease digestion with *Mae*III. The DNA fragments were then treated with calf intestinal phosphatase

(CIP), extracted with phenol-chloroform, and precipitated with ethanol. The purified fragments were first phosphorylated with T4 kinase and [γ-^{32}P]ATP (3,000 Ci/mmol; Amersham Corp., Arlington Heights, Ill.) and were then ligated to each other at 14°C for 2 h with T4 DNA ligase. Both *Ava*I and *Mae*III generated noncomplementary ends that only religated in their original orientations. The products of ligation were digested with restriction enzymes to produce both specific and nonspecific probes. A 347-bp probe that was uniquely labeled within the UPS was generated by *Ava*I digestion of the products of ligation, followed by digestion with CIP, to remove ^{32}P from the 5′ ends. This probe was designated the *Mae*III probe. The *Mae*III probe was used in the experiments for which the results are shown in Fig. 4C, 5, and 6B. A 347-bp probe control was generated by *Mae*III and CIP digestion of the ligation products, which generated a radiolabeled internal *Ava*I site and no intact UPS sequence. This was referred to as the *Ava*I probe. One half of the UPS sequence in this probe was located at either end of the DNA fragment, but neither was labeled with ^{32}P. Alternatively, the correct 347-bp Ad2 MLP fragment was end labeled for use as an additional control. The specificity of protein(s) that bound to the UPS sequence was also examined with ^{32}P introduced at the *Nci*I site (*Nci*I probe) which was at the 3′ border of the UPS core element.

In the second procedure, four synthetic oligonucleotides corresponding to the 5′ and 3′ regions of the UPS on the coding and noncoding strands were used (see Fig. 8B). One member of each pair of oligonucleotides was 5′ end labeled with T4 kinase and [γ-^{32}P]ATP and then ligated to its partner with T4 DNA ligase. This generated single-stranded, full-length coding and noncoding sequences containing ^{32}P within the UPS. The double-stranded DNA probe was then formed by hybridization of the complementary full-length oligonucleotides of the coding and noncoding sequences. This method is particularly useful when specific radiolabeling is desired in binding sites that do not have unique restriction endonuclease sites. The specific introduction of

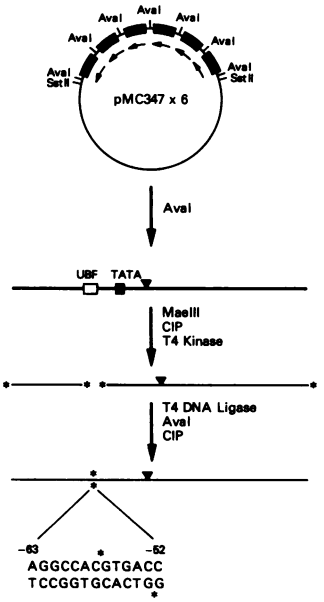

FIG. 1. Construction of the Ad2 MLP, 347-bp, ^{32}P-labeled *Mae*III UPS probe by specific labeling at the *Mae*III restriction endonuclease site.

^{32}P into any phosphodiester bond is readily achieved. This probe was used in the experiment for which the results are shown in Fig. 8.

In the third method, synthetic oligonucleotide templates and primers were used to generate specifically labeled probes. A 20-base primer was annealed to the 3′ end of the corresponding 63-base template (positions −40 to −91) by slow cooling from 65 to 24°C in 2× React 2 buffer (Bethesda Research Laboratories Inc., Gaithersburg, Md.). DNA polymerase I (large fragment) was then used with 10 nmol of [α-^{32}P]TTP (400 Ci/mmol; Amersham) and 2 μmol each of dATP and dGTP, to extend the primer to the first dCTP required within the UPS. After a 10-min incubation at 24°C, 2 μmol each of dCTP and dTTP were added to complete the synthesis of the internally labeled, 63-bp probe. The 63-bp probe (see Fig. 6D) was used in the experiments for which the results are shown in Fig. 6A and C, 7, and 9. Identical strategies were also used when single-stranded bacteriophage M13 clones containing the 347-bp Ad2 MLP sequence in either orientation, and a variety of primers were used.

Oligonucleotide synthesis was performed on a DNA synthesizer (Coder 300; VEGA) by using phosphoramidite chemistry (5). Synthetic oligonucleotides were purified by PICS-high-performance liquid chromatography, as described previously (43).

Labeled probes were purified by 8% PAGE in TBE gels, eluted into 0.5 M $(NH_4)_2CH_3COOH$–10 mM Tris (pH 7.9)–1 mM EDTA, and dialyzed against 10 mM Tris (pH 7.9)–1 mM EDTA before use.

UV cross-linking. Ad2 MLP DNA (positions −138 to +198) uniquely radiolabeled at specific restriction endonuclease sites within the UPS, in sequences flanking the UPS, or both or synthetic oligonucleotide probes containing the UPS were incubated at 24°C for 10 min with 1- to 2-μl fractions of the concentrated Mono S fractions in a total of 20 μl of buffer B [40 mM $(NH_4)_2SO_4$, 1 mM EDTA, 20 mM Tris hydrochloride (pH 7.9), and 1 mM dithiothreitol]. In general, 10 fmol of the ^{32}P-labeled DNA probe was incubated with 1 to 10 μg of protein. The binding reaction was then irradiated with short-wavelength UV light (260 to 280 nm) at 4°C with an average intensity of 600 μW/cm^2. Following the specified time of irradiation, $CaCl_2$ was added to a final concentration of 10 mM, followed by the addition of 50 U of micrococcal nuclease. After a 30-min digestion at 37°C, fractions were resolved by 12% SDS-PAGE and analyzed by autoradiography.

Other procedures. DNase I footprinting, DNA mobility shift analysis, and footprint analysis of DNA shift bands have been described in detail elsewhere (7; Cohen et al., in preparation). Southwest blotting was performed as described previously (1). Protein analysis by SDS-PAGE was performed as described previously (34). G+A ladders used to calibrate DNA footprinting gels were generated by the method of Maxam and Gilbert (28).

Materials. Radiolabeled nucleotides were obtained from Amersham. Restriction endonucleases, T4 polynucleotide kinase, T4 ligase, and Klenow enzyme were purchased from New England BioLabs, Inc. (Beverly, Mass.). Micrococcal nuclease was obtained from Boehringer Mannheim Biochemicals (Indianapolis, Ind.).

RESULTS

Mono S chromatography of HeLa WCE resolves two UPS-binding factors. To identify factors which bind to the UPS of the Ad2 MLP, HeLa WCE was fractionated by cation-

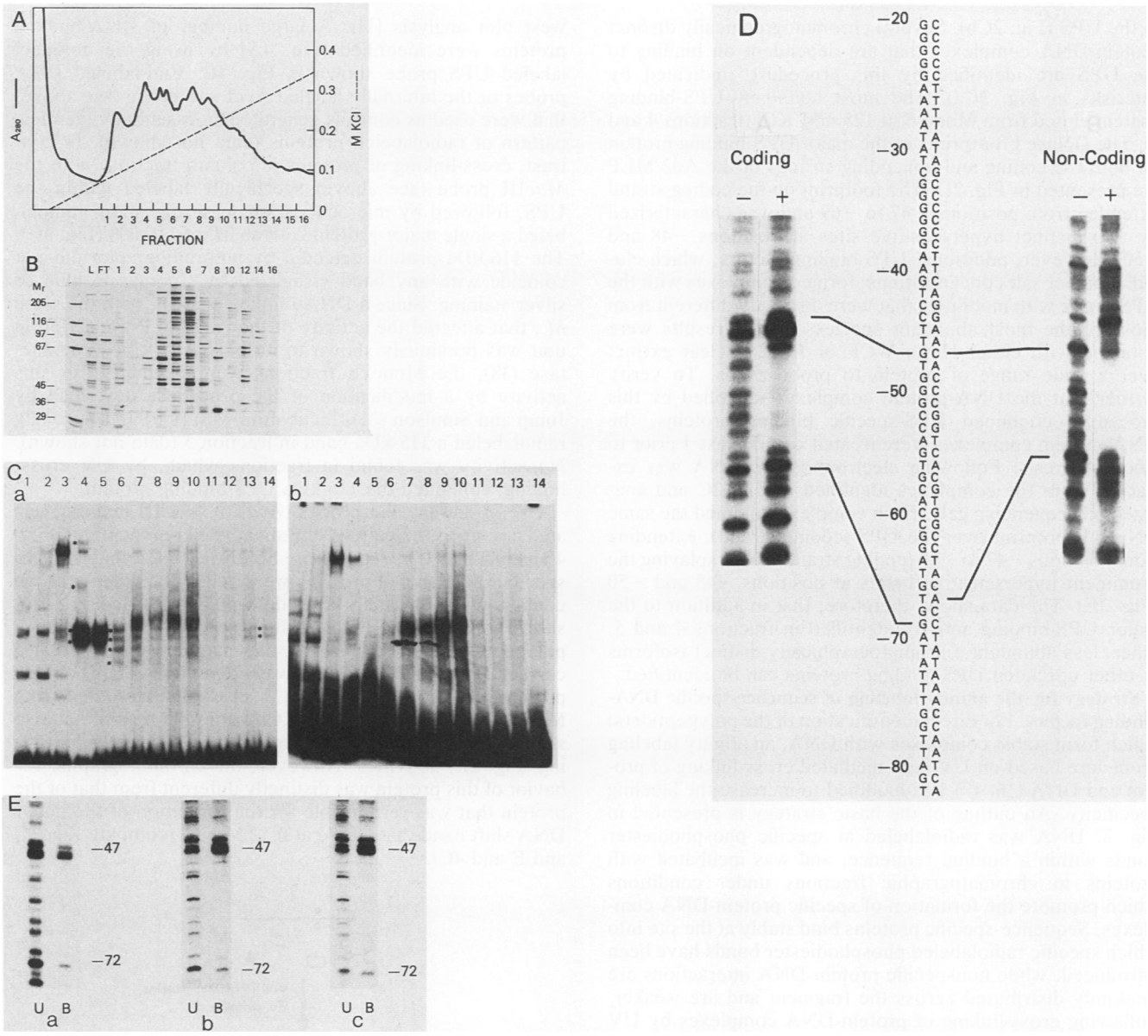

FIG. 2. Chromatography of HeLa WCE on Mono S reveals UPS-binding factor heterogeneity. (A) HeLa WCE (2 ml, 9.6 mg) protein was fractionated by Mono S chromatography in buffer A. Bound proteins were eluted with 45 ml of a linear KCl gradient, from 0 to 0.6 M. The A_{280} and gradient profiles are shown. (B) Coomassie blue-stained 12% SDS-PAGE analysis of the eluted proteins. Abbreviations: L, Load; FT, flow through. (C) DNA mobility shift assay for UPS-binding proteins. Binding reactions (25 μl) contained 1 μl of the Mono S gradient fractions, 1 μg of poly(dI-dC) · (dI-dC), and 0.2 ng of the 5′, ^{32}P-labeled UPS probe (10,000 dpm). The major UPS probe-factor complex was present as a characteristic doublet in fractions 4 and 5. Less abundant UPS-binding protein-DNA complexes were also identified in fractions that eluted at higher salt concentrations (Ca). A duplicate assay was performed in the presence of an unlabeled, 63-bp UPS oligonucleotide competitor (Cb). (D) DNase I footprint analysis of the major UPS-binding activity eluting in Mono S fractions 4 and 5. The position and extent of the UPS-binding factor footprints on both the coding and noncoding strands were mapped with respect to the cap site (+1) of the MLP. (E) DNase I footprint analysis of other UPS-binding protein-DNA complexes identified in panel C. Complexes in fraction 4 (Ea) and the upper two (Eb) and the lower two (Ec) bands in fraction 6 yielded identical footprint patterns. Lanes U and B refer to unbound and bound DNA, respectively. The DNA probe used was the 230-bp *Ava*I-*Dde*I fragment that was 5′ or 3′ end labeled at the *Ava*I site (position −138). Each 25 μl of reaction mixture contained 1 μl of the protein fraction, 1 μg of poly(dI-dC) · (dI-dC) carrier DNA, and 10 fmol of the labeled probe.

exchange chromatography on Mono S resin (Pharmacia). The A_{280} profile and the Coomassie blue-stained polypeptide pattern of proteins, which first eluted with a linear KCl gradient and which were then analyzed by SDS-PAGE, are shown in Fig. 2A and B. Individual fractions were first analyzed for UPS-binding factors by the DNA mobility shift assay described by Fried and Crothers (11). By using a 230-bp end-labeled Ad2 MLP fragment (positions −138 to +98), several distinct DNA-protein complexes were generated (Fig. 2Ca). To identify those complexes that are dependent on specific binding to the UPS, a duplicate mobility shift assay was performed in the presence of a 100-fold molar excess of a synthetic 63-bp DNA fragment that contained the UPS but that lacked other promoter elements (positions −40 to −91). This identifies (by elimination) those protein-DNA complexes that are dependent on sequence-specific binding

to the UPS (Fig. 2Cb). Several chromatographically distinct protein-DNA complexes that are dependent on binding to the UPS are identified by this procedure (indicated by asterisks in Fig. 2Ca). The most abundant UPS-binding protein eluted from Mono S at 125 mM KCl (fractions 4 and 5). The DNase I footprints of the major UPS-binding protein on both the coding and noncoding strands of the Ad2 MLP are presented in Fig. 2D. The footprint on the coding strand extended from positions −47 to −69 and was characterized by two distinct hypersensitive sites at positions −48 and −50. However, additional UPS-binding factors, which eluted at higher salt concentrations, formed complexes with the UPS probe with mobilities that were distinctly different from those of the most abundant species. Similar results were obtained with either HeLa WCE or K562 nuclear extract over a wide range of protein to probe ratios. To verify further that the DNA-protein complexes identified by this procedure contained UPS-specific binding proteins, the DNA-protein complexes were treated with DNase I prior to electrophoresis. Following electrophoresis, DNA was extracted from the complexes identified in Fig. 2C and analyzed on sequencing gels. Each complex contained the same DNase I footprint over the UPS (coding strand), extending from positions −47 to −69 (coding strand) and displaying the prominent hypersensitive bands at positions −48 and −50 (Fig. 2E). The data show, therefore, that in addition to the major UPS-binding activity identified in fractions 4 and 5, other, less abundant, chromatographically distinct isoforms or other unrelated UPS-binding proteins can be identified.

Strategy for the affinity labeling of sequence-specific DNA-binding factors. For direct identification of the polypeptide(s) which form stable complexes with DNA, an affinity labeling procedure based on UV light-mediated cross-linking of protein and DNA (16, 17) was modified to increase the labeling specificity. An outline of the basic strategy is presented in Fig. 3. DNA was radiolabeled at specific phosphodiester bonds within a binding sequence, and was incubated with proteins in chromatographic fractions under conditions which promote the formation of specific protein-DNA complexes. Sequence-specific proteins bind stably at the site into which specific radiolabeled phosphodiester bonds have been introduced, while nonspecific protein-DNA interactions are randomly distributed across the fragment and are weaker. Following cross-linking of protein-DNA complexes by UV irradiation, DNA-protein adducts were digested with micrococcal nuclease. Extensive nuclease digestion left only small oligonucleotides covalently linked to protein. Consequently, radiolabeled oligonucleotides were specifically transferred only to those proteins that were directly in contact with the specific DNA-binding sequence containing ^{32}P-labeled phosphodiester bonds. Nonspecific DNA-binding proteins were not visualized since micrococcal nuclease digestion separates these proteins from ^{32}P-labeled binding sites. Fractions were then analyzed for ^{32}P-labeled proteins by SDS-PAGE. The mobility of the micrococcal nuclease-treated polypeptides was not significantly altered during SDS-PAGE (16), and the molecular weight of the sequence-specific DNA-binding protein was obtained directly by autoradiography.

Affinity radiolabeling identifies a UPS-binding protein with an M_r of 116,000. The DNA affinity labeling procedure was first evaluated for the identification of proteins that bound to an upstream promoter element of the Ad2 MLP (15, 18, 20). Initially, polypeptides in the HeLa WCE fractions resolved by Mono S chromatography were analyzed for UPS-binding proteins by incubating radiolabeled probes with proteins that were transferred to nitrocellulose after SDS-PAGE (Southwest blot analysis [1]). A large number of DNA-binding proteins were identified (Fig. 4A) by using the uniquely labeled UPS probe shown in Fig. 4E. End-labeled DNA probes or the internally labeled *Ava*I site probe (see above) that were used as controls generated an essentially identical pattern of radiolabeled proteins (data not shown). In contrast, cross-linking of proteins in column fractions with the *Mae*III probe (see above) specifically labeled within the UPS, followed by micrococcal nuclease digestion, radiolabeled a single major protein with an M_r of 116,000 (Fig. 4C). The 116-kDa protein detected by autoradiography did not coincide with any band visualized by Coomassie blue or silver staining. Since a DNA-binding protein, with the same M_r, that affected the activity of the Ad2 MLP transcription unit was previously shown to be poly(ADP)ribose synthetase (38), the Mono S fractions were examined for this activity by a modification of the procedure described by Jump and Smulson (21). Incubation with [^{32}P]NAD strongly radiolabeled a 115-kDa band in fraction 5 (data not shown). No activity was found in fractions which, by UV cross-linking, contained the 116-kDa UPS-binding protein.

Use of end-labeled probes (*Ava*I or *Mae*III ends) generated a nonspecific pattern of radiolabeled polypeptides (Fig. 4B) similar to that seen in the Southwest blot analysis. No specific radiolabeled proteins were generated when a probe containing the internally labeled *Ava*I site was used (data not shown). The ability to radiolabel a single, low-abundance polypeptide in a sequence-specific manner suggests that a correct UPS-binding factor was identified. The 116-kDa protein identified by specific UV cross-linking eluted from Mono S from 0.25 to 0.32 M KCl within the region that also showed DNA mobility shift (Fig. 2C) and DNase I footprinting (Fig. 2E) activities. However, the chromatographic behavior of this protein was distinctly different from that of the protein that was responsible for the formation of the major DNA shift band that eluted at 0.12 M KCl (compare Fig. 2C and E and 4C).

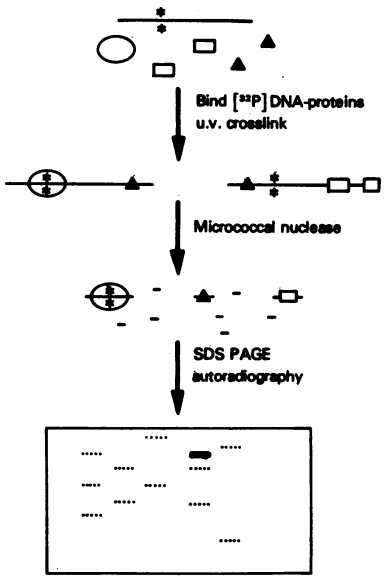

FIG. 3. Strategy for the direct identification of DNA-binding proteins. UV cross-linking specificity was achieved by the incorporation of ^{32}P exclusively within the specific binding sequence and by the separation of nonspecific protein-DNA cross-links from these ^{32}P-labeled phosphodiester bonds by extensive micrococcal nuclease digestion.

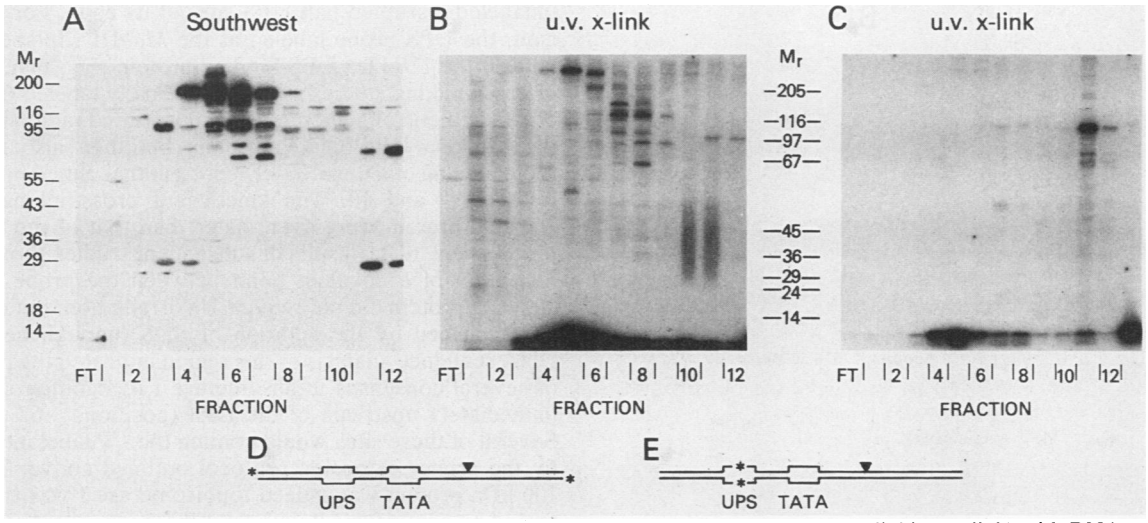

FIG. 4. Comparison of DNA-binding proteins identified by Southwest blot analysis and UV cross-linking (x-link) with DNA probes that were radiolabeled within the UPS or 5′ end labeled. (A) Proteins displayed in Fig. 1B were electrophoretically transferred to nitrocellulose. Following partial renaturation in buffer A, the blot was probed with [^{32}P]Ad2 MLP DNA (positions −138 to +196) that was radiolabeled as indicated in panel D or E. Identical results were obtained with probe that was labeled specifically within the UPS (E) or 5′ end labeled (D), indicating the inability of identifying sequence-specific DNA-binding proteins by this procedure. (B) Nonspecific DNA-binding proteins were identified by the 5′-end-labeled probe (D) by using the UV cross-linking procedure. (C) A single 116-kDa DNA-binding protein was identified with the specific UPS probe labeled internally at the *Mae*III restriction endonuclease site within the UPS (E). The positions of ^{32}P within these probes are indicated with asterisks (D and E). The duration of UV irradiation was 5 min. FT, Flow through ▼, transcription start site.

Detection of a second UPS-binding factor requires extended UV cross-linking times. To visualize the protein that eluted at a low salt concentration from Mono S and that was responsible for the major DNA mobility shift band, several parameters of the cross-linking protocol were examined. The first important variable was found to be the duration of UV irradiation. With extended irradiation times, a second 45-kDa protein that eluted in fractions 4 and 5 was identified with a *Mae*III-labeled UPS probe. Details of the purification of the 45-kDa UPS-binding factor will be reported elsewhere (Cohen et al., in preparation). The time course of DNA cross-linking to the 45- and 116-kDa UPS-binding proteins was distinctly different. While cross-linking of both the 116- and 45-kDa proteins was absolutely dependent on UV irradiation, the half time of cross-linking to the former was approximately 2 min, while cross-linking to the latter proceeded linearly for 60 min. Simultaneous detection of both UPS-binding proteins in Mono S fractions with the *Mae*III probe therefore requires a much longer period of UV irradiation than was initially used. In addition, the amount of nonspecific competitor DNA required for cross-linking specificity varied with both proteins. UV cross-linking of the 45-kDa UPS protein was enhanced by poly(dI-dC) · poly (dI-dC) at early stages of its purification, while cross-linking of the 116-kDa UPS protein was not.

The second important variable was the amount of ^{32}P-labeled probe that was required to detect factors in partially purified fractions. An analysis of UV cross-linking to the 116- and 45-kDa UPS-binding proteins as a function of probe concentration is presented in Fig. 5. Mono S fractions containing either the 45- or the 116-kDa proteins were combined in a ratio which made the band intensities of both proteins comparable on autoradiography. The mixture of UPS-binding proteins was then irradiated for 30 min. At limiting concentrations of ^{32}P-labeled probe (2 fmol), the 45-kDa protein, but not the 116-kDa protein, was seen. At higher concentrations of probe, both proteins were readily visualized. The data illustrate, therefore, that at limiting concentrations of probe, not all sequence-specific DNA-binding proteins may be detected.

Sequence specificity of the affinity labeling procedure. To demonstrate the sequence-specific nature of the proteins which we identified using the affinity labeling procedure, we performed several types of control experiments. In the first of these, ^{32}P-containing bonds were introduced at different sites within the UPS or in flanking regions just outside the UPS. One such probe was produced by using a synthetic 63-nucleotide oligonucleotide of the noncoding strand (positions −40 to −91) which was annealed to a complementary 20-nucleotide primer (positions −71 to −91). Primer extension with the Klenow enzyme in the presence of [^{32}P]TTP, dGTP, and dATP but in the absence of dCTP limited the incorporation of deoxyribonucleotides to a region 10 bases 3′

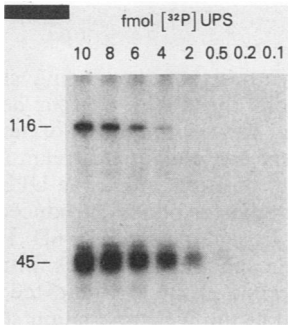

FIG. 5. Effect of probe concentration on UV cross-linking to the 116- and 45-kDa UPS-binding proteins. Mono S fractions 5 and 12 were combined and incubated with increasing amounts of the specific internally labeled 347-bp UPS probe for 10 min. UV cross-linking analysis was then performed by using a 30-min UV exposure period. The amounts of 45- and 116-kDa proteins used were chosen such that bands with comparable intensities could be produced in the presence of saturating levels of probe. The relative band intensities therefore do not reflect their relative abundances in unfractionated HeLa WCE.

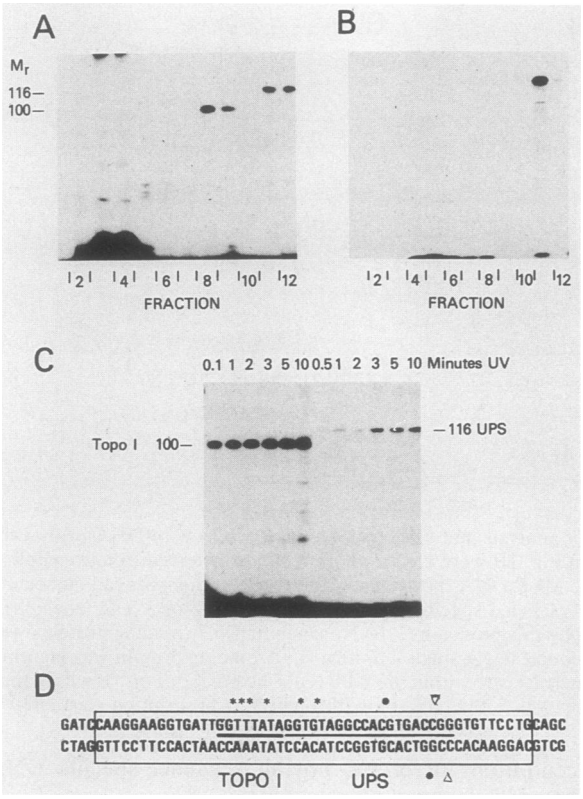

FIG. 6. The specificity of UV cross-linking is highly dependent on the ^{32}P-labeled phosphodiester bond position. ^{32}P-labeled phosphodiester bonds were introduced within the UPS at the *Mae*III site (●) or at the *Nci*I site just 3′ to the UPS core sequence (△) or within the UPS and its immediate 5′-flanking region by primer extension (*). Unique cross-linking of the 116-kDa UPS-binding protein was obtained only with the *Mae*III site-labeled probe. (A) No cross-linking was observed with the *Nci*I site-labeled probe (data not shown). Probe labeled by primer extension identified not only the 116-kDa protein but also a 110-kDa protein that eluted from Mono S at a lower KCl concentration. (B) Several consensus topoisomerase I (TOPO I)-binding sites in the immediate flanking region 5′ to the UPS were identified by sequence analysis (indicated by the heavy line in panel D; the lighter line in panel D indicates the UPS core sequence). The kinetics of UV cross-linking of the 100- and 116-kDa proteins to Ad2 MLP probe labeled at the asterisks is shown in panel C.

to the primer. Synthesis of the coding strand was then completed by a cold chase with all four deoxyribonucleotides. This strategy limited the incorporation of ^{32}P to the region beginning six nucleotides upstream of the UPS and ending at the fourth position within the UPS (Fig. 6D). The ^{32}P-labeled phosphodiester bonds introduced by this procedure are indicated by asterisks in Fig. 6D. For comparison, the ^{32}P-labeled phosphodiester bonds introduced at the *Mae*III site in a second probe are indicated by two dots. A third probe labeled at the *Nci*I restriction site just 3′ to the UPS was also constructed, and the ^{32}P-labeled bonds introduced in this manner are indicated by triangles.

By using the *Nci*I probe, no DNA-binding proteins in any of the fractions eluted from Mono S were specifically identified by the UV affinity labeling procedure (data not shown). This was not surprising since the *Nci*I site lies at the 3′ border of the UPS core sequence. This was also the result which we obtained earlier with the permuted Ad2 MLP 347-bp probe containing an internally labeled *Ava*I site with unlabeled disrupted half UPS sites at its ends. For comparison, the UPS probe labeled at the *Mae*III site specifically labeled the 116-kDa UPS-binding protein (Fig. 6B). Results of cross-linking studies with the internally labeled 63-nucleotide synthetic oligomer again demonstrated radiolabeling of the 116-kDa UPS-binding protein but then also revealed radiolabeling of a new 100-kDa protein that eluted in Mono S fractions 9 and 10. The kinetics of cross-linking of the 100-kDa protein were even faster than that of the 116-kDa protein (Fig. 6C). Results of subsequent studies showed that formation of a covalent bond between the probe and the 100-kDa protein did not require UV irradiation, but could be accomplished by the addition of SDS (unpublished data). DNA sequence analysis of this region indicated the presence of several consensus topoisomerase I recognition sites (39) immediately upstream of the UPS (positions −67 to −73). Several of these sites would contain the ^{32}P label introduced by the primer extension protocol outlined above. That the 100-kDa protein was indeed topoisomerase I was then confirmed by copurifying it with an ATP-independent relaxation activity on supercoiled plasmids (S. Sturm, personal communication). Thus, the exact site of the ^{32}P-labeled phosphodiester bond within the DNA probe determines the specificity obtained by this affinity transfer procedure.

In the second type of control experiment, we employed unlabeled competitor fragments to decrease specific cross-linking of UPS-binding proteins to the UPS. For these studies we used the internally radiolabeled, 63-bp fragment described above and protein fractions containing the 45-kDa, UPS-binding protein identified in Mono S fractions 4 and 5 (Fig. 2C). As competitor fragments we used either a cloned 230-bp Ad2 MLP fragment that extended from positions −138 to +98 and that contained the UPS or a cloned 149-bp fragment that contained an intact Ad2 MLP TATAA sequence but that lacked the entire UPS. The extent of UV cross-linking in the absence of competitor is shown in Fig. 7 (lane 5). With increasing amounts of the 230-bp DNA fragment, the extent of UV cross-linking was decreased. No competition was observed, however, with the 149-bp fragment which lacked the UPS. The specificity of the UV affinity transfer procedure was thus confirmed by two independent techniques.

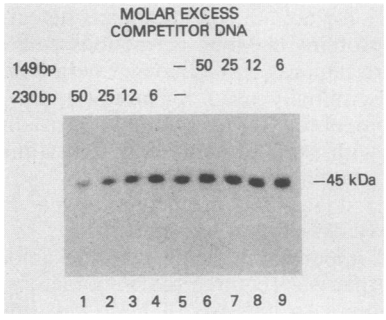

FIG. 7. Competition analysis of the specificity of UV cross-linking of the 45-kDa protein to the UPS. The 45-kDa UPS-binding protein was incubated with 10 fmol of the internally labeled 63-bp UPS probe and increasing amounts of unlabeled Ad2 MLP fragments containing (positions −138 to +98) or not containing (positions −37 to +28) the UPS. Following 10 min of incubation, UV cross-linking was performed for 30 min. With competitor DNA containing the UPS, UV cross-linking was essentially eliminated at a 50-fold molar excess. A similar inhibition did not occur with the nonspecific competitor DNA. Lane numbers are indicated at the bottom of the gel.

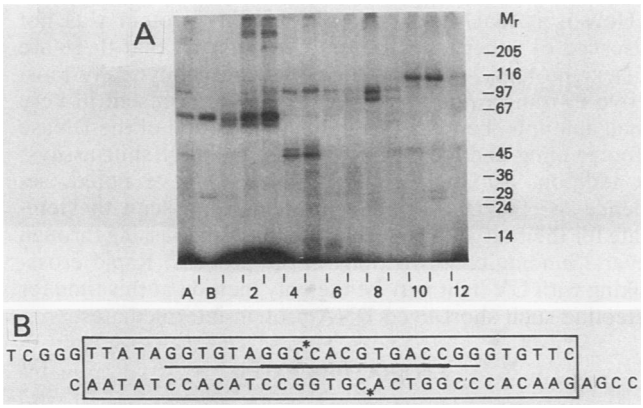

FIG. 8. Specificity of UV cross-linking is affected by probe size. (A) UV cross-linking to proteins in Mono S fractions was performed with a 38-bp synthetic oligonucleotide probe. The placement of ^{32}P within the core sequence of the UPS is indicated by asterisks (B). UV cross-linking analysis was performed over a 30-min irradiation period. Both the 116- and 45-kDa UPS-binding proteins were identified in fractions 10 and 11 and 4 and 5, respectively (fraction numbers are indicated at the bottom of the gel). Several other unidentified proteins become radiolabeled, however, in contrast to the results obtained with the longer 347- and 63-bp probes (Fig. 4C and 6B, respectively).

Size also affects probe specificity. To examine the effect of probe size on the specificity of UV cross-linking, an internally labeled 38-bp probe containing ^{32}P within the UPS was constructed by ligation and subsequent hybridization of four synthetic oligonucleotides (Fig. 8B). Use of this shorter probe to identify UPS-binding proteins eluting from Mono S by affinity cross-linking resulted in the identification of both the 45- and 116-kDa UPS-binding proteins (Fig. 8A, fractions 4 and 5, and lanes 10 and 11, respectively). A pattern of nonspecific protein labeling was also present, however, that was similar to that generated with any end-labeled DNA fragment. This limited the usefulness of such short probes in the initial identification of specific DNA-binding proteins. Presumably, the nonspecific background arises from end-binding proteins which extend along the DNA oligomer sufficiently to allow UV cross-linking, which was a distance of 17 nucleotides in this particular probe. Once the identity of the affinity-labeled proteins is established, however, such probes can be used to monitor purification of the specific DNA-binding proteins. Alternatively, UV cross-linking can be performed in the presence of higher levels of nonspecific unlabeled DNA competitors which increase the signal to noise ratio in this and other assay systems that are dependent on sequence-specific protein-DNA interactions.

DISCUSSION

Deletion and mutation analysis of the Ad2 MLP identifies a *cis*-acting element upstream of the TATAA box which regulates the rate of expression of this transcription unit (15, 18, 42, 20). In vitro transcription activity of the Ad2 MLP exhibited a 10- to 20-fold dependence on this sequence. This UPS contains an imperfect, centrally located palindrome, GGCCACGTGACC. By DNase I footprint analysis, it was determined that the UPS spans this palindrome and extends from positions −47 to −69, with two hypersensitive sites on the coding strand located at positions −48 and −50. In several laboratories a 42- to 55-kDa protein has been purified which can generate the DNase I footprint, and this protein has been designated as the upstream element factor (30), major late transcription factor (6), upstream binding (UB) factor (35), or upstream stimulatory (US) factor (37). In a recent report (6), it has been indicated that purification to apparent homogeneity was achieved. A major problem that we encountered, however, was an apparent loss of transcriptional activity with increasing purification of this transcription factor (R. B. Cohen and S. Garfinkel, unpublished data). Similar observations have been made by other investigators during the purification of UPS-specific DNA-binding proteins. Carthew et al. (4) noted a requirement for a fraction that eluted at a high salt concentration from cation-exchange resins in their fractionated assay system as the major late transcription factor was increasingly purified. Moncollin et al. (30) also reported an increasing loss of transcriptional activity with the highly purified upstream element factor. Several other transcription factors, including Sp1 (22), NF-1/CTF (8, 19, 22), heat shock transcription factor (40), and IID (32), have also been reported to exist in several discrete molecular weight forms (33). Such heterogeneity is likely to result from purification of distinct but related species. Since several DNA-binding proteins which regulate transcription unit activity have separate DNA sequence recognition and effector domains (3, 23, 24, 29), we considered the possibility that the partial assays used to purify the UPS factor might result in the purification of only one of these domains. Most partial assays that are currently used for factor purification rely on DNA-binding activity or effects on DNA structure, which alter accessibility to enzymatic or chemical modification, rather than activity. The extent of DNA-binding activity, however, need not correlate directly with functional activity during transcription. Specifically, a loss of catalytic activity by modification of the functional domain could occur during purification, independent of any alteration of the DNA recognition domain. To study the native form of these factors, we sought to develop an analytical procedure by which, in crude or partially fractionated extracts, those proteins that are responsible for sequence-specific DNA-binding activity could be identified directly. If possible, the method could be employed early in a fractionation procedure to identify the number and native M_rs of DNA-binding proteins that recognize a specific DNA sequence and to devise additional strategies for purification of these proteins.

Photochemical cross-linking of proteins to nucleic acid was explored as a method by which sequence-specific DNA-binding proteins could be identified directly. Photochemical cross-linking was first used to study binding of *lac* repressor to *lac* operator DNA (25, 31) and histone binding to DNA and chromatin (27). Subsequently, cross-linking was used to identify those subunits of a highly purified RNA polymerase that were in direct contact with DNA (16). More recently, a significant improvement was achieved through the application of laser technology, which increased the efficiency and rapidity of cross-linking (17). In all of these studies, however, uniformly labeled probes were used. It was not possible in these studies, therefore, to be certain that DNA-protein cross-linking occurred exclusively at specific binding sites. In addition, in all of these studies either highly purified or very abundant proteins were used; this limited the number of radiolabeled bands that were generated. Since formation of covalent bonds between DNA and protein is confined to points of intimate contact, we reasoned that specific transfer of label from DNA to protein would occur if radiolabel was placed only within DNA sequences that were recognized by their specific cognate-binding protein. Proteins that cross-

linked to the specifically radiolabeled DNA at nonspecific binding sites would not be visualized by autoradiography, since only short, nonradioactive oligonucleotides would remain cross-linked to them following nuclease digestion. In addition, the molecular weight of the specific DNA-binding protein could then be determined by SDS-PAGE, since the mobility of proteins linked to short nucleic acid oligomers or monomers is not significantly altered (16).

Specific cross-linking of proteins bound to the UPS of the Ad2 MLP was achieved by limiting the introduction of ^{32}P to phosphodiester bonds within the UPS. By using 260- to 280-nm UV irradiation, cross-linking occurred without any requirement for the incorporation of bromodeoxyuridine into the DNA probe (17). Indeed, the only requirement for successful UV-mediated cross-linking appears to be the presence of radiolabeled pyrimidines within the binding site. This is advantageous for several reasons. First, single-strand breaks at bromodeoxyuridine sites in the DNA probes resulting from UV irradiation are avoided, thus minimizing the background of nonspecific proteins identified by this procedure. Second, sequence-specific DNA-binding proteins that recognize sites that do not contain thymidine can still be identified. It is noteworthy that when thymidine-containing probes were used, as in this study, the slow kinetics of UV cross-linking of the 45-kDa UPS-binding factor were comparable to those in the recent report of Chodosh et al. (6), in which bromodeoxyuridine-substituted probes were used.

The most important feature of the specific probes used in this study was that radiolabel was introduced solely within the binding sequence. In addition, we found that it is important to introduce radiolabel at an internal site to avoid labeling the nonspecific proteins that bind to fragment ends. When low-abundance, DNA-binding proteins are studied, the use of uniformly labeled probes results in a large increase in the number of radiolabeled polypeptides generated by the affinity cross-linking procedure. Since these probes are generally easier to make and are still able to identify the specific binding protein, in addition to the nonspecific bands, they may nevertheless be useful once the specific band is identified unambiguously. They cannot be used, however, in an initial screening procedure in crude extracts for rare sequence-specific DNA-binding proteins.

Protein detection by UV cross-linking with site-specific labeled probes may also provide a useful technique for estimating the pool sizes of DNA-binding proteins. The two methods used to date, DNase I footprinting and gel electrophoresis mobility shift assays, are poorly suited to this purpose. DNase I footprinting, which is perhaps the superior of the two techniques for pool size estimates, requires the presence of saturating amounts of protein to generate detectable protection from DNase I. This requirement limits its applicability to DNA-binding proteins which are present in subsaturating amounts in crude or partially purified extracts. The gel electrophoresis mobility shift assay is probably more sensitive to the presence of limited amounts of protein, but this assay is also subject to interference by nonspecific binding proteins. UV cross-linking, when performed in the manner that we have described here, is sensitive enough to detect proteins beneath the threshold of the DNase I footprint assay. If radiolabel is placed only in the binding site, the nonspecific DNA-protein interactions complicating interpretation of mobility shift assays are largely eliminated. Careful titrations to ensure probe excess (Fig. 5) and determination of the intensity and/or duration of UV irradiation are necessary, however.

How is it that the larger UPS-binding protein was not observed by other investigators? We suspect that there are nuclear proteins, perhaps those that are catalytically most active in transcription initiation, which are present in very small amounts, beneath the limits of detection of the DNase I footprinting and gel electrophoresis mobility shift assays. In addition, as Hockensmith et al. (17) have noted, sequence-specific DNA-binding proteins may scan the template for their target by a sliding mechanism, passing through several binding conformations in the process. Rapid cross-linking with UV light may be the only method at this time for detecting such short-lived DNA-protein intermediates.

LITERATURE CITED

1. **Bowen, B., J. Steinberg, U. K. Laemmli, and H. Weintraub.** 1980. The detection of DNA binding proteins by protein blotting. Nucleic Acids Res. **8**:1–20.
2. **Breathnach, R., and P. A. Chambon.** 1981. Organization and expression of eukaryotic split genes. Annu. Rev. Biochem. **50**:349–383.
3. **Brent, R., and M. Ptashne.** 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell **43**:729–736.
4. **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.
5. **Caruthers, M. H.** 1985. Gene synthesis machines. DNA chemistry and its uses. Science **230**:281–285.
6. **Chodosh, L. A., R. W. Carthew, and P. Sharp.** 1986. A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol. Cell. Biol. **6**:4723–4733.
7. **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.
8. **Diffley, J. F. X., and B. Stillman.** 1986. Purification of a cellular, double-stranded DNA-binding protein required for initiation of adenovirus DNA replication by using a rapid filter-binding assay. Mol. Cell. Biol. **6**:1363–1373.
9. **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.
10. **Dynan, W. S., and R. Tjian.** 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature (London) **316**:774–778.
11. **Fried, M. G., and D. M. Crothers.** 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9**:6505–6525.
12. **Galas, D., and A. Schmitz.** 1978. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. **5**:3157–3171.
13. **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.
14. **Heinz, N., and R. G. Roeder.** 1982. Transcription of eukaryotic genes in soluble cell-free systems, p. 57–89. *In* J. K. Setlow and A. Hollaeunder (ed.) Genetic engineering, vol. 4. Plenum Publishing Corp., New York.
15. **Hen, R., 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.
16. **Hillel, Z., and C.-W. Wu.** 1978. Photochemical cross-linking studies on the interaction of Escherichia coli RNA polymerase with T7 DNA. Biochemistry **17**:2954–2961.
17. **Hockensmith, J. W., W. L. Kubasek, W. R. Vorachek, and P. H. von Hippel.** 1986. Laser crosslinking of nucleic acids to proteins: methodology and first applications to the phage T4 DNA replication system. J. Biol. Chem. **261**:3512–3518.

18. **Hu, S.-L., and J. L. Manley.** 1981. DNA sequences required for initiation of transcription in vitro from the major late promoter of adenovirus 2. Proc. Natl. Acad. Sci. USA **78**:820–824.
19. **Jones, K. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelley, and R. Tjian.** 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell **48**:79–89.
20. **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.
21. **Jump, D. B., and M. Smulson.** 1980. Purification and characterization of the major nonhistone protein acceptor for poly(adenosine diphosphate ribose) in HeLa cell nuclei. Biochemistry **19**:1024–1030.
22. **Kadonaga, J. T., and R. Tjian.** 1986. Affinity purification of sequence-specific DNA binding proteins. Proc. Natl. Acad. Sci. USA **83**:5889–5893.
23. **Kameshita, I., Z. Matsuda, T. Taniguchi, and Y. Shizuta.** 1984. Poly(ADP-ribose) synthetase separation and identification of three proteolytic fragments as the substrate binding domain, the DNA-binding domain and the automodification domain. J. Biol. Chem. **259**:4770–4776.
24. **Keegan, L., G. Gill, and M. Ptashne.** 1986. Separation of DNA binding from transcription activating function of a eukaryotic regulatory protein. Science **231**:699–704.
25. **Lin, S. Y., and A. D. Riggs.** 1974. Photochemical attachment of lac repressor to bromodeoxyuridine-substituted lac operator by ultraviolet radiation. Proc. Natl. Acad. Sci. USA **71**:947–951.
26. **Manley, J. L., A. Fire, A. Caro, P. A. Sharp, and M. L. Getter.** 1980. DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA **77**:3855–3859.
27. **Martinson, H. G., M. D. Shetlar, and B. J. MaCarthy.** 1976. Histone-histone interactions within chromatin. Crosslinking studies using ultraviolet light. Biochemistry **15**:2002–2007.
28. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
29. **McKnight, S., and R. Tjian.** 1986. Transcriptional selectivity of viral genes in mammalian cells. Cell **46**:795–805.
30. **Moncollin, V., N. G. Miyamoto, X. M. Zheng, and J. M. Egly.** 1986. Purification of a factor for the upstream element of the adenovirus-2 major late promoter. EMBO J. **5**:2577–2584.
31. **Ogata, R., and W. Gilbert.** 1977. Contacts between the lac repressor and thymines in the lac operator. Proc. Natl. Acad. Sci. USA **74**:4973–4977.
32. **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.
33. **Reinberg, D., M. Horikoshi, and R. G. Roeder.** 1987. Factors involved in specific transcription in mammalian RNA polymerase II. J. Biol. Chem. **262**:3322–3330.
34. **Safer, B., L. Yang, H. E. Tolunay, and W. F. Anderson.** 1985. Isolation of stable preinitiation, initiation and elongation complexes from RNA polymerase II directed transcription. Proc. Natl. Acad. Sci. USA **82**:2632–2636.
35. **Sawadogo, M., and R. G. Roeder.** 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43**:165–175.
36. **Shenk, T.** 1981. Transcriptional control regions: nucleotide sequence requirements for initiation by RNA polymerase II and III. Curr. Top. Microbiol. Immunol. **93**:25–46.
37. **Shi, X.-P., R. Lee, and R. Weinman.** 1986. Protein factors binding independently to two different regions of the adenovirus 2 major late promoter. Nucleic Acids Res. **14**:3729–3744.
38. **Slattery, E., J. D. Dignam, T. Matsui, and R. G. Roeder.** 1983. Purification and analysis of a factor which suppresses nick-induced transcription by RNA polymerase II and its identity with poly(ADP-ribose) polymerase. J. Biol. Chem. **258**:5955–5959.
39. **Vosberg, J.-P.** 1985. DNA topoisomerases: enzymes that control DNA conformation. Curr. Top. Microbiol. Immunol. **114**: 19–102.
40. **Wiederrecht, G., D. J. Shuey, W. A. Kibbe, and C. S. Parker.** 1987. The Saccharomyces and Drosophila heat shock transcription factors are identical in size and DNA binding properties. Cell **48**:507–515.
41. **Wu, C.** 1985. An exonuclease protection assay reveals heat shock element and TATA box DNA binding proteins in crude nuclear extracts. Nature (London) **317**:84–87.
42. **Yu, Y. T., and J. L. Manley.** 1984. Generation and functional analysis for base substitution mutants of the adenovirus 2 major late promoter. Nucleic Acids Res. **12**:9309–9321.
43. **Zon, G., and J. A. Thompson.** 1986. A review of high-performance liquid chromatography in nucleic acids research. II. Isolation, purification and analysis of oligodeoxyribonucleotides. Biochromatography **1**:22–32.