Identification of sequence-specific DNA-binding factors by label transfer: application to the adenovirus-2 major late promoter

L.G.Nikolaev, B.O.Glotov, A.V.Belyavsky, S.A.Grachev and A.V.Levin

Institute of Molecular Biology, USSR Academy of Sciences, Moscow 117984, USSR

Received September 8, 1987; Revised and Accepted December 14, 1987

ABSTRACT

A method of affinity labelling proteins specifically associated with DNA target sequences is proposed. The method utilizes covalent UV-crosslinking of proteins to highly labelled DNA (e.g. in crude cell or nuclear extracts) followed by degradation of the DNA to short oligonucleotides. Proteins selectively labelled by attached residual oligonucleotides are readily amenable to molecular mass determination. Using this approach, we have characterized a HeLa polypeptide specifically bound to a short segment of the adenovirus-2 major late promoter (Ad2 MLP). A molecular mass value (~ 51 kD) and precise location of the crosslinking site(s) of the protein within the MLP (~ 55 with respect to the cap site) were determined.

INTRODUCTION

In recent years a number of investigators have been studying the mechanisms of transcriptional control of eukaryotic protein-coding genes (class II genes). Control and promoter regions of these genes are known to contain distinct DNA elements, usually located upstream and in cis with respect to the transcription initiation site, which determine the position of this site and also modulate the effectiveness of transcription (1,2).

There is evidence that the effects of the upstream elements are mediated by sequence-specific proteins called transcription factors (for reviews see 3,4). Isolation and characterization of transcription factors, including determination of their abundance and activity throughout cell growth and development, are indispensable for understanding the principles of transcription control.

In spite of efforts few proteins related to eukaryotic transcription factors have been well characterized up to now (5-9), primarily due to the very low content of these factors in the cell. Transcription factors have in most cases been analysed by purification using affinity or standard chromatography as well as by experiments revealing protection of control and/or promoter elements from DNaseI attack (10) or chemical modification (11).

In the present work we demonstrate an alternative approach

which allows one to measure molecular mass values and thus identify sequence-specific DNA-binding proteins even in crude protein extracts. The approach is based on affinity labelling of proteins selectively bound to DNA target sequences of interest. In brief, the procedure is as follows.

Cloned and uniformly labelled ^{32}P DNA of the region of interest is incubated with whole cell, nuclear or other protein extracts. In the presence of a large excess of unlabelled foreign competitor DNA the promoter DNA binds predominantly sequence-specific proteins. The incubation mixture is then irradiated with ultraviolet (UV) light, resulting in the formation of covalently linked DNA-protein complexes (DNP's) which can be electrophoretically separated and recovered from gels in the form of individual DNP's. DNA within each DNP is enzymatically or chemically degraded to short oligonucleotides. The residual radioactively labelled oligonucleotides crosslinked to proteins selectively mark the polypeptide chains bound to the promoter DNA (12-15). An overwhelming proportion of other (non-specific) DNA-binding proteins associate with the unlabelled competitor DNA. The proteins affinity labelled in this way are readily amenable to molecular mass determination using SDS-PAGE.

This approach was used to study DNA-protein interactions in mixtures of the Ad2 MLP region with a HeLa whole cell protein extract (WCE). Transcription from this promoter has been extensively studied in recent years in vivo and in vitro. The Ad2 MLP was shown to contain at least two transcriptional control regions. The first one (-21 to -31 with respect to the cap site) determines the correct placement of RNA polymerase II on the DNA (9,16), while the second one (-53 to -62) seems to be necessary for effective transcription in vivo and in vitro (6,17,18). Mutations in the latter region reduce the effectiveness of transcription by a factor of 3-5 (18).

The affinity labelling procedure described above allowed us to identify a HeLa polypeptide specifically binding the Ad2 MLP, to measure its molecular mass (~51 kD) and to map the site of its covalent crosslinking to the MLP (-55).

MATERIALS AND METHODS

Restriction endonucleases, alkaline phosphatase of *E. coli*, T4 DNA-ligase, T4 DNA polymerase, Klenow fragment of *E. coli* DNA polymerase I, exonuclease III of *E. coli*, 5'-exonuclease of λ phage were obtained from "Ferment" (USSR). DNaseI (Worthington), staphylococcal nuclease and poly d(AT) (Boehringer-Mannheim), pronase (Calbiochem), deoxynucleoside triphosphates, HEPES and *E. coli* tRNA (Sigma) were used. [α - ^{32}P]dCTP with a specific activity of 3000 Ci/mmol was obtained from "Izotop" (USSR).

The 17-nt M13 sequencing primer was prepared by Dr B.K. Chernov. 5'-Bromo-2'-deoxyuridine triphosphate (BrdU) was obtained by bromination of 2'-deoxyuridine triphosphate (sodium salt, Sigma) and purified by ion-exchange chromatography on DEAE-cellulose.

Plasmids, phages and DNA fragments

A -250 to +32 Ad2 MLP fragment was cut out of plasmid p ϕ 4 (17) with restriction endonucleases HindIII and XhoI and reclon-

ed between the polylinker HindIII and SalI sites of phages M13mp18 and M13mp19 (19) using standard protocols (20).

Unlabelled MLP for competition experiments was the HindIII-XhoI fragment of p ϕ 4 isolated by electrophoresis on an agarose gel (21).

The SalI-BglII fragment of 3'-structural portion of the *Drosophila* hsp 70 gene of the insert 132 E3 (22) was electrophoretically purified and used as non-promoter DNA. A whole cell extract of HeLa S3 grown in suspension culture was prepared as previously described (23).

The band shift assay in polyacrylamide gels for MLP DNA mixed with a WCE HeLa was carried out according to (24).

Preparation of labelled and BrdU-substituted MLP

1 μ g M13 DNA with an insert of the MLP (annealed with the 17-nt sequencing primer) was dissolved in 20 μ l of a buffer containing 50 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 0.1 mM dithiothreitol (DTT), 50 μ g/ml BSA, 25 μ M of each dATP, dGTP, BrdU (triphosphate), and 150 μ Ci [α -³²P]dCTP. After the addition of 2 units of the Klenow fragment the mixture was incubated for 1 h at 20°C, then dCTP was added to a final concentration of 50 μ M, and the incubation was continued for additional 20 min. To stop the reaction, 5 M sodium chloride was added to a concentration of 50 mM, and the enzyme was inactivated by heating for 15 min at 65°C. The DNA in the mixture was cut with restriction endonucleases HindIII and BamHI (using 5 units of each) for 30 min at 37°C. The digestion was stopped by the addition of 1/10 volume of 2% SDS, 0.2 M Na₂EDTA, 50% glycerol, 0.025% bromophenol blue, and the resulting mixture was applied to a non-denaturing 12% polyacrylamide gel in a Tris-acetate buffer system (20).

An autoradiographically localized MLP band was excised from the gel. MLP DNA was eluted from the gel pieces as described (25), precipitated with ethanol, washed and redissolved in a minimum volume of TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM Na₂EDTA).

As a result, double stranded MLP DNA was obtained in which one of the strands (referred to as the bl-strand) was labelled with ³²P cytidine residues and contained BrdU residues substituted for thymidine. Bl-coding and bl-noncoding strands were synthesized using inserts in M13mp18 and M13mp19, respectively.

Substitution of thymidine residues by BrdU results in an increase of the yield of DNA-protein UV-crosslinks by an order of magnitude.

For the treatment with λ 5'-exonuclease, the MLP with bl-coding strand was synthesized as above but after heat inactivation of the enzyme the DNA was first cleaved with HindIII only, the sticky end being filled in with freshly added Klenow fragment and cold nucleoside triphosphates. The enzyme was again heat inactivated, and the MLP with a blunt HindIII end was excised with BamHI and isolated as described above.

UV-crosslinking and isolation of crosslinked DNP's

Labelled MLP (0.5 ng) was mixed with a WCE HeLa (1-5 μ g of protein) and poly d(AT) (1.5 μ g) and incubated for 30 min at 30°C in 15 μ l of 15 mM HEPES-NaOH, pH 7.9, 60 mM KCl, 7 mM MgCl₂, 0.3 mM Na₂EDTA, 1.2 mM DTT, 10% glycerol. In a preparative version the amount of MLP was increased to 10-20 ng with a proportional increase in WCE, poly d(AT), and the reaction volume.

The incubated mixture was transferred to a quartz cell with an optical path length of 1 mm and irradiated with UV-light at 254 nm at an intensity of 1.2×10^{16} photon/min cm^2 (1.6 W/m^2) for 3-10 min. Crosslinked DNP's were enriched by separation from the bulk of uncrosslinked DNA using sorption on a phenol/buffer interface (26) after addition of an equal volume of 2xBEP buffer (0.1 M Tris-HCl, pH 8.0, 1 M NaCl, 2% SDS) and 15-min incubation at 50°C. The enriched DNP's were ethanol precipitated, washed, and redissolved in 25 μl of 50 mM Tris-HCl buffer, pH 6.8, 10% glycerol, 5% SDS, 3% 2-mercaptoethanol. The solution was then heated for 30 min at 50°C and applied to a 5% SDS-polyacrylamide gel according to (27).

Nuclease digestion of DNP's (DNA to protein label transfer)

Autoradiographically localized DNP bands were excised from the gel and individual DNP's were eluted as described (25) using a Mg-free buffer containing 15 $\mu\text{g/ml}$ tRNA. The eluted DNP's were precipitated with ethanol, washed and redissolved each in 20-30 μl of TE buffer. To the solutions 1/10 volume of 0.5 M Tris-HCl, pH 8.0, 50 mM MgCl_2 , phenylmethylsulphonyl fluoride to a concentration of 0.5 mM and 20 units of DNaseI were added. The reaction was conducted for 2-30 min at 37°C and terminated by addition of SDS, glycerol and bromophenol blue to concentrations of 2, 20 and 0.0025%, respectively. After heating for 5 min at 100°C digested DNP's were applied to a 15% SDS-polyacrylamide gel according to (27).

Digestion with staphylococcal nuclease was done for 5 min at 37°C in 50 mM Tris-HCl buffer, pH 7.5, 5 mM CaCl_2 in the presence of 30 units of the enzyme.

Polyacrylamide gel electrophoresis

PAGE was performed in slabs (17x17x0.1 cm). For two-dimensional electrophoresis of the "band shift-DNP" type a first dimension gel strip ("band shift") was excised from the gel, immersed and incubated in 50 mM Tris-HCl, pH 6.8, 1% SDS, 2% 2-mercaptoethanol for 30 min at 65°C and then put on the top of a 5% SDS-polyacrylamide gel (27) for electrophoresis in the second dimension.

For two-dimensional PAGE of the type "DNP-protein" a first dimension gel strip (crosslinked DNP) was cut out of the gel, immersed in 50 mM Tris-HCl buffer, pH 8.0, 5 mM MgCl_2 and incubated for 30 min at 37°C and then for additional 30 min in the presence of 30 units/ml of DNaseI. The strip was next incubated in 50 mM Tris-HCl, pH 6.8, 1% SDS, 2% 2-mercaptoethanol for 10 min at 100°C and put on the top of a 12% SDS-polyacrylamide gel for separation in the second dimension.

Denaturing PAGE of DNA was done on 6% polyacrylamide gels containing 7 M urea in a Tris-borate buffer as described (25).

One- and two-dimensional analytical gels were dried (after removal of urea, if necessary) and autoradiographed.

Exonuclease footprinting

Individual DNP's prepared as described for DNaseI treatment were digested in 60 μl of a buffer for T4 DNA polymerase (33 mM Tris-OAc, pH 7.9, 10 mM MgOAc , 66 mM KOAc, 0.5 mM DTT, 100 $\mu\text{g/ml}$ BSA) containing 3 units of T4 DNA-polymerase and 4 units of ExoIII. The reaction mixture was incubated for 30 min at 37°C (limit-digest conditions), and the reaction was terminated by an addition of 40 μl of a stop-buffer (1% SDS, 40 mM Na_2EDTA , 0.5 mg/ml sheared calf thymus DNA). This was followed by a pronase

treatment (0.5 mg/ml, 1.5 h, 37°C), phenol deproteinization, ethanol precipitation, redissolution in formamide and heat denaturation of DNA (20,25). Finally, the DNA products of exonuclease digestion were electrophoresed on a 6% denaturing polyacrylamide gel.

Digestion of DNP's with 5'-Exo was done for 1 h at 37°C in 30 µl of a buffer (7 mM glycine, pH 9.4, 3 mM MgCl₂, 3 mM 2-mercaptoethanol) containing 5 units of the enzyme. To stop the reaction, 40 µl of the stop-buffer and 30 µl of distilled water were added. Further treatment and PAGE of the DNA products were as described for 3'-exonuclease.

Restriction endonuclease "footprinting"

Individual DNP's prepared as above were dissolved in 60 µl of a low salt restriction endonuclease buffer, pH 7.5 (20), containing 0.5 mg/ml BSA, and digested with 12 units of restriction endonuclease MboII for 2 h at 37°C. 50 µg BSA, 40 µl of distilled water and 100 µl of 2xBEP buffer were then added, and the procedure of DNP enrichment on a phenol/buffer interface (see above) was repeated 3 times. To ether-extracted aqueous phases containing enriched DNP's (~50 µl) 2 volumes of distilled water were added followed by pronase treatment (0.3 mg/ml, 2 h, 37°C). Further treatment and PAGE of the resulting DNA was as described for the exonuclease footprinting.

The products of the digestion not retained on the interface (protein-free DNA) were also collected, deproteinized, ethanol precipitated and similarly analysed on 6% denaturing polyacrylamide gels.

RESULTS

DNA-protein complexes in non-denaturing gels ("the band shift assay")

To optimize the conditions of DNA-protein binding we used the band shift assay based on the reduction of electrophoretic mobility of a DNA fragment when it is complexed with protein molecules (28,29). The MLP fragment (Fig. 1) with the coding strand substituted with bromodeoxyuridine (BrdU) and radioactively labelled (bl-coding strand, see MATERIALS AND METHODS)

Fig. 1. A DNA insert used as the Ad2 MLP region. The initial MLP region of the Ad2 genome (-253 to +32) is flanked by a synthetic HindIII linker and a BamHI-SalI stretch of the M13 phage vector polylinker. Nucleotide coordinates are given for a coding strand. +1 corresponds to the first nucleotide of the Ad2 MLP RNA transcript. Arrows show the most probable sites of pf51 UV-crosslinking to the coding strand. MboII restriction endonuclease break points within the coding strand are shown.

Fig. 2. Nondenaturing electrophoresis of MLP-protein complexes ("the band shift assay"). 4% polyacrylamide gel at low ionic strength. Lane 1, bl-MLP (1 ng) mixed with a HeLa WCE (6 μ g of protein) and poly d(AT) (2 μ g). Lanes 2-5, the same as 1, but after addition of 2, 5, 10 and 50 ng of unlabelled MLP, respectively. Lanes 6 and 7, the same as 1 but after addition of 10 and 50 ng of unlabelled non-promoter DNA. A and B designate the major and one of the minor MLP-protein complexes, respectively.

was mixed with a HeLa WCE and unlabelled poly d(AT) as competitor DNA. The mixture was electrophoresed on a non-denaturing polyacrylamide gel to give the autoradiograph depicted in Fig.2. Excess unlabelled MLP (Fig. 2, lanes 1-5) or non-promoter DNA of similar length (lanes 6 and 7) was used to test the specificity of DNA-protein binding.

In association with proteins, the MLP fragment migrates in the gel as a major band A accompanied by a number of minor bands including a poorly seen band B, migrating close to band A. The addition of excess unlabelled MLP results in a decrease in the intensity of band A. The effective competition of unlabelled MLP for binding of protein(s) contained in the band A DNP implies the sequence-specific binding of the MLP.

Under the same conditions the intensities of the minor band B and the other minor bands are considerably less sensitive, indicating that these represent non-specific protein-DNA interactions. Alternatively, the proteins in these DNP's may be very abundant in the protein extract.

Similar titration with non-promoter DNA fragments, as expected, fails to change the intensity of any DNP bands.

Crosslinked DNA-protein complexes

A mixture of labelled MLP with a HeLa WCE and poly d(AT), similar to those used in the band shift assay, was irradiated

Fig. 3. Electrophoresis of UV-crosslinked MLP-protein complexes (DNP) on a 5% SDS-polyacrylamide gel. Lane 1, UV-irradiated mixture of MLP with bl-coding strand (18 ng), a HeLa WCE (100 μ g of protein) and poly d(AT)(30 μ g). Lane 2, the same as lane 1 but for MLP with bl-noncoding strand. Major bands of covalently crosslinked DNP's are numbered.

with UV-light followed by the removal of excess uncrosslinked DNA (only a small fraction of the DNA is crosslinked to proteins under the conditions used, see Discussion). Enriched covalently crosslinked DNP's were separated by SDS-PAGE. The autoradiographs of the gels for MLP containing bl-coding and bl-noncoding strands are shown in Fig. 3. The lane 1 for the MLP with bl-coding strand contains three prominent bands (DNP-1, -2 and -3) and several minor bands. For the MLP with bl-noncoding strand there can be also seen a number of dominant (DNP-4, -5 and -6) and minor bands.

The DNP bands are absent when UV-irradiation is omitted and disappear after pronase treatment (data not shown). The radioactive bands in Fig. 3 thus seem to represent the DNP species formed due to covalent crosslinking of the proteins to labelled MLP fragments under UV-irradiation.

The DNP bands representing proteins crosslinked to either bl-coding or bl-noncoding strands are qualitatively very much alike but quantitatively differ considerably. In particular, the major bands of DNP's on lane 1 correspond to the minor bands on lane 2 (or are absent) and vice versa. This indicates

Fig. 4. Electrophoresis of MLP-protein complexes (DNP's) UV-cross-linked in the presence of unlabelled MLP (competition experiments). 5% SDS-polyacrylamide gel.

A, MLP with bl-coding strand (1.5 ng) mixed with a HeLa WCE (11 μ g of protein) and poly d(AT)(3 μ g), with addition of 0, 5 and 30 ng of unlabelled MLP (lanes 1-3, respectively).

B, lanes 4-6, the same as A, lanes 1-3, respectively, but for MLP with bl-noncoding strand. Major DNP bands are numbered.

unequal effectiveness of MLP-protein crosslinking by UV-radiation for MLP's with different bl-strands.

It should be noted that for labelled but not BrdU substituted MLP fragments the patterns analogous to those in Fig. 3 practically coincide (data not shown).

The specificity of MLP-protein UV-crosslinking was verified by adding excess unlabelled MLP or non-promoter DNA fragments prior to UV-irradiation, as was done in the band shift assay. Unlabelled MLP competes effectively with the labelled MLP for binding of the DNP-1 protein (Fig. 4 A), as shown by a decrease and then disappearance of the DNP-1 band following the stepwise addition of cold MLP. The same titration does not affect the other major bands (DNP-2 and DNP-3 for the bl-coding strand, and DNP-4, -5 and -6 for the bl-noncoding strand) or the minor bands. A 30-fold excess of the non-promoter DNA fragment also has no effect on intensities of any DNP bands (data not shown).

We note that, as judged from the competition experiments, the only protein specifically binding MLP is that contained in the DNP-1. Other MLP-protein adducts seem to incorporate

Fig. 5. Two-dimensional PAGE "band shift-crosslinked DNP". MLP with bl-coding strand (2 ng) was mixed with a HeLa WCE (3.5 μ g of protein) and poly d(AT) (3 μ g), UV-irradiated and electrophoresed on a low ionic strength 4% polyacrylamide gel (first dimension). 5% SDS-polyacrylamide gel (second dimension).

proteins which are non-specific and/or highly abundant in the protein extracts.

Correlation between the proteins binding MLP in the band shift assay and those crosslinked to MLP by UV-light

To correlate the DNP bands in the band shift assay (Fig. 2) with patterns of UV-crosslinked MLP-protein complexes (Fig. 3) a UV-irradiated mixture was analysed on two-dimensional gels. The first dimension was run in the conditions of the band shift assay. Except for a slight broadening of the bands, the pattern did not differ from that without UV-irradiation. In the second dimension the material was separated by SDS-PAGE. An autoradiograph of a resulting two-dimensional gel for MLP with bl-coding strand (Fig. 5) reveals that bands A and B in the band shift assay comigrate in the first dimension with crosslinked DNP-1 and DNP-3, respectively. In spite of considerable difference in electrophoretic mobilities on SDS-polyacrylamide gels, DNP-1 and DNP-3 almost (but not entirely) comigrate under non-denaturing conditions.

The identity of the proteins involved in the formation of the DNP-1 and the DNP of band A is supported by similar sensitivity of these DNP's to titration with unlabelled MLP. The unexpected ratio of intensities for the DNP-1 and DNP-3 spots in Fig. 5 with respect to those of bands A and B in the first dimension can be explained by a higher yield of crosslinking of band B protein to DNA.

An analogous result was also obtained for MLP with bl-noncoding

Fig. 6. Determination of molecular mass value of a protein incorporated in DNP-1 (DNA to protein label transfer). 15% SDS-polyacrylamide gel. Lane 1, ^{125}J -protein length markers (chymotrypsinogen - 25 kD, catalase monomer - 60 kD, and crosslinked catalase dimer and trimer - 120 and 180 kD). Lane 2, undigested DNP-1; lane 3, DNP-1 digested with DNaseI; lane 4, the same as 3 but after pronase treatment (1 mg/ml, 37°C, 30 min); lane 5, DNP-1 digested with staphylococcal nuclease. An arrow indicates the protein band (pf51).

strand but in this case the DNP-1 spot was much less intense (not shown).

Identification of the DNP-1 protein by its molecular mass

To determine the molecular mass of the protein specifically binding MLP contained in DNP-1 we carried out a preparative MLP-protein UV-crosslinking (for an initial amount of DNA of 15-20 ng) followed by electrophoretic separation of the crosslinked DNP's. The DNP-1 band was cut out of the gel, and DNP-1 eluted from the gel pieces was treated with DNaseI or staphylococcal nuclease, the products being then analysed by SDS-PAGE. An autoradiograph of the resulting gel is presented in Fig. 6. Prior to nuclease digestion DNP-1 does not enter a 15% polyacrylamide gel (Fig. 6, lane 2) but in the course of the digestion there appears a discrete radioactive band (lane 3) sensitive to pronase (lane 4). The band represents the protein of DNP-1 retaining radioactive label in the form of the covalently attached short single-stranded oligonucleotide.

The average chain length of residual oligonucleotides was estimated by measuring the fraction of radioactivity bound to the protein after nuclease digestion. For DNaseI and staphylococcal nuclease this length was calculated to be 5-6 and 1 nucleotide (nt), respectively. In both cases the electrophoretic mobilities of the protein were similar suggesting a limited influence of short oligonucleotides on protein mobility during SDS-PAGE (lanes

Fig. 7. Two-dimensional PAGE "DNP-proteins". First dimension, UV-crosslinked MLP-protein complexes separated on a 5% SDS-polyacrylamide gel (conditions are given in the legend to Fig. 2). Second dimension, 12% SDS-polyacrylamide gel. A and B, MLP with bl-coding and bl-noncoding strands, respectively. Protein length markers as in Fig. 5.

3 and 5). An apparent molecular mass value for the protein of DNP-1, measured in three independent experiments, was found to be 50.6 ± 1.2 kD (Fig. 6, lane 3). For the sake of convenience, this protein will be further referred to as pf 51 (protein factor), and the procedure described as DNA to protein label transfer.

Characterization of proteins of non-specific DNP's

To obtain the total molecular mass spectrum of all the proteins crosslinked to the MLP we used two-dimensional SDS-PAGE where UV-crosslinked DNP's were separated in the first dimension, and the proteins after the label transfer were separated in the second dimension. A gel of this kind for the MLP with bl-coding strand is shown in Fig. 7. Apart from the dominating spot 1 of pf51 there are protein spots corresponding to apparent molecular mass values of 100 and 130 kD from the zone of DNP-2, 230-240 kD from DNP-3, as well as several other spots of 230-240 kD proteins.

As seen from the pattern in Fig. 7A, electrophoretic mobilities of DNP's in the first dimension are in general inversely proportional to the molecular mass values of their associated proteins, although there are noticeable deviations. For example, DNP corresponding to a 130-kD protein (spot 3) migrates slightly faster than that having a 100-kD protein (spot 2).

A protein spectrum on an analogous two-dimensional gel for the MLP with bl-noncoding strand (Fig. 7B) is different from that

Fig. 8. Identification of pf51 crosslinking sites on the MLP coding strand (exonuclease footprinting). 6% denaturing polyacrylamide gel. A, 3'-exonuclease footprinting. Lane 1, control MLP (UV-irradiated under normal conditions but not crosslinked to proteins) digested with ExoIII+3'-ExoT4; lane 2, DNP-1 digested with ExoIII+3'-ExoT4; lane 3, pBR 322 MspI restriction fragments (length marker, nt). B, 5'-exonuclease footprinting. Lane 1, control MLP; lane 2, control MLP digested with 5'-Exo; lane 3, length marker (nt); lane 4, DNP-1 digested with 5'-Exo. Bands of exonuclease-resistant DNA are indicated by arrows.

for bl-coding strand. In particular, the pf51 spot is practically absent, in agreement with asymmetric crosslinking of this protein to complementary bl-strands of MLP DNA. DNP-4 contains a protein with a molecular mass value of ~150 kD, DNP-5 of ~130 and 230-240 kD, and DNP-6 of 230-240 kD.

Location of proteins on DNA (footprinting)

To precisely localize the sites of protein crosslinking to DNA in various DNP's, the DNP's were digested with 3'- and 5'-exonucleases. The exonuclease footprinting relies on a supposition that a nucleotide with a covalently crosslinked protein chain is an obstacle for exonucleases processing along DNA.

In the footprinting with 5'-exonuclease of λ phage (5'-Exo) only the bl-coding strand of the MLP could serve as substrate due to selective filling-in of a HindIII sticky end prior to UV-irradiation.

For 3'-5' single strand digestion we took advantage of a combined action of exonuclease III from E.coli (ExoIII) and 3'-exonuclease activity of T4 phage DNA-polymerase (3'-ExoT4). The mixture of the enzymes overcomes non-specific DNA "stop-sites"

Fig. 9. Restriction endonuclease "footprinting".

MLP with bl-coding strand; 6% denaturing polyacrylamide gel. Lanes 1-3, DNA-products of DNP-1, -2 and -3, respectively, digested with MboII, not retained on the phenol/buffer interface, i.e. initially free of protein. Lanes 4-6, DNA-products of DNP-1, -2 and -3, respectively, digested with MboII, retained on the phenol/buffer interface, i.e. initially bound to proteins. Lane 7, DNA single strand length marker: pBR 322 MspI restriction fragments, nt.

(barriers for exonucleases) much more effectively than each enzyme taken separately. The "stop-sites" can be readily seen in control digests of free MLP (Fig. 8A, lane 1).

Fig. 8A presents a typical separation of deproteinized DNA-products of DNP-1 exhaustively digested with an ExoIII + 3'-ExoT4 mixture by denaturing PAGE (lane 2). There is one major band of 92 nt and two minor bands of ~26 and ~18 nt of exonuclease-resistant DNA absent from similarly digested control MLP preparations (UV-irradiated in normal mixtures but not crosslinked to proteins).

A digestion of the same DNP-1 with 5'-Exo, that is from the opposite side of bl-coding strand, reveals a single resistant band of ~210 nt "complementary" to the 92 nt band left by 3'-5' digestion (the single strand full length of MLP DNA is 303 nt) and absent from control digests (Fig. 8B).

Statistical analysis of exonuclease-resistant DNA for three independent experiments shows that the major site of pf51 crosslinking to MLP coding strand is mapped at position -92 ± 2 nt from the 5'-terminal nucleotide or at -55 ± 2 nt from the cap site (29) (Fig. 1). Taking into account preferential crosslinking of proteins to BrdU residues it would be reasonable to assume that the exact site of pf51 crosslinking to the MLP coding strand is a BrdU residue at positions -53 or -58 (or both of them) (Fig. 1)

Due to the lower efficiency of UV-crosslinking of pf51 to the MLP bl-noncoding strand, a similar analysis for this strand was impossible.

The existence of minor pf51 crosslinking sites at positions +13 and +21 with respect to the cap site, inside the structural part of the gene, was not confirmed by digestion with 5'-Exo. However, it cannot be excluded that expected "complementary" exonuclease resistant DNA fragments (290 and 282 nt) are obscured by non-specific bands migrating within the same region of the gel (Fig. 8B, lane 2). Anyway, the question of the existence of these minor crosslinking sites remains open.

Exonuclease footprinting of "non-specific" DNP-2 and DNP-3 (both with bl-coding and bl-noncoding strands) does not reveal any bands other than those present in control MLP digests (data not shown). This finding could be accounted for by a multiplicity of protein crosslinking sites on the DNA of DNP-2 and DNP-3 and the lack of any preferential crosslinking site.

To check this possibility, DNP-2 and DNP-3 were partially digested with restriction endonuclease MboII under conditions when the site at -104 was split to a much greater extent than that at -4 (see the map in Fig. 1). As a result, over 90% of the DNA fell apart into two closely sized fragments (162 and 141 nt along the coding strand). Free DNA was removed from the digests by phenol extraction (see Materials and methods), and the enriched DNP's were deproteinized and analysed by denaturing PAGE (Fig. 9, lanes 2 and 3). It can be seen that both halves of DNP-2 and DNP-3 are retained on a phenol/buffer interface with a ratio close to 1:1. Control digests of free MLP DNA are not appreciably retained on the interface (data not shown). This indicates that DNA-protein crosslinking sites are more or less evenly distributed between the two halves of DNP-2 and DNP-3 DNA, in accord with the existence of multiple crosslinking sites.

Surprisingly, an application of the "restriction endonuclease footprinting" to DNP-1 also detects molecules of pf51 crosslinked to both halves of MLP DNA although in this case the contacts with the right half, containing the preferential crosslinking site, predominate (Fig. 9, lanes 1 and 4). Insofar as the probability of simultaneous crosslinking of two protein molecules to the same MLP fragment is extremely low, it appears that apart from the site of pf51 at -55, the MLP also contains other sites of lower affinity. If there were few minor sites and they were strictly localized within the MLP, all of them could probably be revealed by exonuclease footprinting but this is evidently not the case. Therefore the minor sites appear to be dispersed.

It is conceivable that UV-irradiation can fix on DNA the protein molecules involved in one-dimensional diffusion (and therefore localized randomly). Further studies are needed to clarify this possibility unambiguously.

DISCUSSION

The experiments reported here clearly demonstrate that UV-fixation of proteins on highly labelled DNA followed by the transfer of the label to these proteins allows one to identify them by molecular mass determination even at a rather low yield of UV-crosslinking ($< 0.5\%$). Under the conditions used this value is close to a maximum one since attempts to further increase the

yield using longer irradiation times encountered undesirable increase in UV-damaging of the DNA. On the other hand, low efficiency of crosslinking makes simultaneous crosslinking of two or more protein molecules to the same DNA fragment highly improbable, simplifying later analysis.

The conditions chosen for the band shift assay and UV-crosslinking of proteins to the MLP correspond to those of the *in vitro* transcription reaction (23, and our unpublished results) except for the addition of a large excess of competitor poly d(AT) and the lack of nucleoside triphosphates. The pattern of DNP bands for the MLP in the band shift assay (Fig. 2) and their sensitivity to the titration with unlabelled MLP closely resemble those described earlier for the similar Ad2 MLP fragment (24).

Markedly different effectiveness of UV-crosslinking of the proteins to bl-coding and bl-noncoding strands of MLP DNA (Fig. 3) can be explained both by preferential interaction of the protein molecules with only one of the two strands and by asymmetric distribution of BrdU residues within the complementary strands. An effect of BrdU distribution is evident from the identity of DNP electrophoresis patterns for different MLP strands which are radioactively labelled but not substituted by BrdU. This does not, however, rule out selective interaction of proteins with only one of two complementary MLP strands.

The molecular mass value of the sequence-specific protein pf51 (~ 51 kD) is close to an earlier predicted value of 55 ± 8 kD for a similar, if not identical, transcription factor USF partially purified from HeLa protein extracts (7). However, in the work of these authors direct assignment of the protein bands visible on the gel to USF is difficult.

We have also detected several proteins incorporated in non-specific MLP-protein complexes and measured their apparent molecular mass values. None of them have a molecular mass value below 50 kD. Earlier work has described the partial purification of a 43-kD STF factor present in a HeLa WCE which presumably binds to the TATA-box region of the Ad2 MLP (9). The lack of STF-like proteins in our tests could be explained in several ways. Perhaps most pertinent is the question of whether all of the proteins bound to DNA can be fixed by UV-irradiation. Taking into account a pronounced dependence of the yield of UV-fixed DNP's on the nature of proteins (as revealed by the differing intensities of DNP bands in the band shift assay and those of the corresponding crosslinked DNP's) and on the distribution of BrdU residues in DNA strands, the answer should be rather negative. In other words, some MLP-binding protein species probably can avoid covalent crosslinking while other species will be readily crosslinked to DNA.

Exonuclease footprinting of pf51 covalently bound to the MLP allowed us to localize the site of its fixation at -55 ± 2 nt from the transcription start site. Since an exonuclease can be expected to stop at some short distance from the sites of polypeptide chain attachment to DNA (probably 0-3 nt), it seems reasonable to assume that pf51 is actually crosslinked to a single nucleotide. This nucleotide is most probably one of the BrdU residues at positions -53 or -58. Both the residues map within an Ad2 MLP region previously identified as a binding site of transcription factors USF (-63 to -52)(6) and UEF (-67 to -50)(7), suggesting a similarity or identity of those factors to pf51.

The described affinity labelling of proteins is a versatile method which can be applied to analysis of proteins recognizing any DNA sequences of interest (promoters, enhancers etc.).

ACKNOWLEDGEMENTS

The authors are indebted to Drs K.K. Ebralidze, G.N. Enikolopov, A.M. Kolchinsky, G.S. Monastyrskaya and S.G. Arsenyan for valuable discussions and helpful advice. Our thanks are also due to Dr B.K. Chernov for synthesis of the sequencing primer, to Dr R.L. Turetskaya for her help in cultivation of HeLa cells, and to Dr M. Pifer for critical reading of the manuscript.

Note

After this work has been sent to the editor, we obtained a paper (30) very closely tied in with the subject of the present communication. The authors used an analogous approach to characterize specific proteins bound to the Ad2 MLP.

REFERENCES

1. Shenk, T. (1981) Current Topics in Microbiology and Immunology 93, 25-46
2. Heintz, N. and Roeder, R.G. (1982) In Setlow, J.K. and Hollander, A. (eds), Genetic Engineering, Plenum Press, New York, vol. 4, pp. 57-89
3. Dynan, W.S. and Tjian, R. (1985) Nature 316, 774-778
4. Ptashne M. (1986) Nature 322, 697-701
5. Dynan, W.S. and Tjian, R. (1983) Cell 32, 669-680
6. Sawadogo, M. and Roeder, R.G. (1985) Cell 43, 165-175
7. Moncollin, V. and Miyamoto, N.G. (1986) EMBO J. 5, 2577-2584
8. Wiederrecht, G., Shuey, D.J., Kibbe, W.A. and Parker, C.S. (1987) Cell 48, 507-515
9. Egly, J.M., Miyamoto, N.G., Moncollin, V. and Chambon P. (1984) EMBO J. 3, 2367-2371
10. Galas, D.J. and Schmitz, A. (1978) Nucl. Acids Res. 5, 3157-3171
11. Gilbert, W., Maxam, A. and Mirzabekov, A. (1976) In Kjeldgaard, N.O. and Maaloe, O. (eds), Control of Ribosome Synthesis, Alfred Benson Symp. 9, Academic Press, New York, pp. 139-148
12. Hellel, Z. and Wu, C.-W. (1978) Biochemistry 17, 2956-2968
13. Chatterjee, P.K., Vayda, M.E. and Flint, S.G. (1986) J. Mol. Biol. 188, 23-27
14. Ebralidze, K.K., Tuneev, V.M., Mel'nikova, A.F., Gogvadze, E.G. and Mirzabekov, A.D. (1982) Dokl. Akad. Nauk SSSR 265, 1272-1275 (in Russian)
15. Hilton M.D., Whitely H.R. (1985) J. Biol. Chem. 260, 8121-8127
16. Hu, S.-L. and Manley, J.L. (1981) Proc. Natl. Acad. Sci. USA 78, 820-824
17. Jove, R. and Manley, J.L. (1984) J. Biol. Chem. 259, 8513-8521
18. Yu, Y.-T. and Manley, J.L. (1984) Nucl. Acids Res. 12, 9303-9321
19. Yanish-Perron, C., Viera, J. and Messing, J. (1985) Gene 33, 103-119

-
20. Davis, L.G., Dibner, M.D. and Batley, J.F. (1986) *Basic Methods in Molecular Biology*, Elsevier Science Publishers, New York
 21. Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) *Anal. Biochem.* 112, 295-298
 22. Karch, F., Torok, I. and Tissieres, A. (1981) *J. Mol. Biol.* 148, 219-230
 23. Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Gelfand, M.L. (1980) *Proc. Natl. Acad. Sci. USA* 77, 3855-3859
 24. Cartew, R.W., Chodosh, L.A. and Sharp, P.A. (1985) *Cell* 43, 439-448
 25. Maxam, A.M. and Gilbert, W. (1980) In *Methods of Enzymology*, Vol. 65, part 1, pp. 499-560
 26. Djerbash'yan, A.R., Vashakidze, R.P., Karpov, V.L., Kolchinsky, A.M. and Mirzabekov, A.D. (1987) *Molek. Biologiya*, submitted (in Russian)
 27. Laemmli, U.K. and Favre, M. (1973) *J. Mol. Biol.* 80, 601-613
 28. Fried, M. and Crothers, D.M. (1981) *Nucl. Acids Res.* 9, 6505-6525
 29. Baker, C.C. and Ziff, E.B. (1981) *Cold Spring Harbor Symp. Quant. Biol.* 44, 415-428
 30. Chodosh L.A., Cartew R.W., Sharp P.A. (1986) *Mol. Cell. Biol.* 6, 4723-4733