

Methods for the Analysis of DNA-Protein Interactions

*Matthew J. Guille and G. Geoff Kneale**

Abstract

The interaction of proteins with DNA is a central theme of molecular biology. In this article, we review some of the principal techniques currently used for the identification and characterization of DNA binding proteins, and for investigation of the molecular interactions that are responsible for the recognition of specific DNA sequences.

Index Entries: DNA-binding proteins; transcription factors; DNA modification; DNA footprinting.

1. Introduction

Increased understanding of DNA-protein interactions has become a major goal for biologists because it is clear that these interactions play a central role in a wide range of cellular processes. Examples include transcription regulation, DNA replication, genetic recombination, DNA repair, and the restriction and modification of DNA. The types of investigation carried out on DNA-protein interactions fall into two main categories. In the first type, a DNA sequence has been shown to have function, for example, by deletion of a promoter region of a gene, and information is needed about the protein and its interaction with this sequence in order to understand its role in transcription control. In the second type, the protein is already well-characterized and the investigation concentrates on the precise molecular nature of the contacts between the protein and DNA, and the possible conformational changes in either component that might result, often as a prelude to fuller structural studies of the complex.

In this article, we review both of these types of investigation, which have substantial areas of overlap. The techniques described permit the identification of DNA-binding proteins in cell extracts, and give information on the size of the protein(s) involved, the location of the DNA-binding site, the strength and specificity of binding, the identities of

individual groups or specific bases involved in binding, and the effects of protein binding on gross conformation or local structure of DNA.

The recognition of DNA sequences by proteins is a complex phenomenon, involving specific hydrogen bonding contacts to the DNA bases and/or interactions with the sugar-phosphate backbone. The latter interactions can also be highly specific because of the sequence-dependent structural polymorphism of DNA. In addition, intercalation of planar aromatic amino acid side chains between the DNA bases can occur, most notably with single-stranded DNA binding proteins. Furthermore, when bound, many DNA-binding proteins induce drastic structural changes in the DNA that are essential to their function. The advantages of using complementary methods to investigate the complexities of protein-DNA interactions will be clear, as each has advantages and disadvantages for specific tasks. For this reason, it is unwise to rely on a single technique when interpreting the results of such investigations.

2. Identification and Preliminary Characterization of DNA-Binding Proteins

2.1. The Electrophoretic Mobility Shift Assay (EMSA) in Crude Protein Extracts

If initial studies (perhaps from functional or genetic analysis) have suggested that a particular

* Author to whom all correspondence and reprint requests should be addressed. Division of Molecular and Cell Biology, School of Biological Science, University of Portsmouth, Portsmouth PO1 2DT, UK, e-mail: geoff.kneale@port.ac.uk.

DNA sequence is recognized by a protein within the cell, EMSA is often undertaken using this DNA sequence as a probe in crude cell extract. There are a number of reasons for this:

1. To confirm that one or more proteins interact with this region of DNA specifically;
2. To test the distribution of this complex between different cell-types or growth states; and
3. To identify which subcellular compartment the protein is in.

The EMSA (also known as a gel retardation assay, or “gel-shift” assay) is relatively straightforward (1,2). A radioactively labeled DNA duplex, either a short restriction fragment (25–75 bp) or a synthetic oligonucleotide, is incubated with whole-cell, cytoplasmic or nuclear extract in the presence of competitors to prevent nonspecific binding. (Poly[dI.dC], tRNA, and salmon sperm DNA may all be used.) DNA–protein complexes are then allowed to form and the free DNA and complexes are resolved on a low ionic strength, low percentage polyacrylamide gel. The complexes and free DNA are then visualized by autoradiography (Fig. 1). In these assays of crude extracts, a number of complexes are often formed with only a subset specific for the DNA sequence under test. Specificity of binding is assessed by inclusion of an excess of unlabeled competitor DNA in the binding reaction. An excess of unlabeled probe competes for specific complexes, but point-mutant versions of the probe or unrelated DNA sequences do not (Fig. 1).

The simplicity of this assay, its high sensitivity, and the fact that many protein samples may be assayed simultaneously make it the usual method of choice for assaying DNA–protein complexes from a number of cell-types; for example, when testing the tissue distribution of a transcription factor or when a time course of binding activity is needed (3). The small amount of protein needed for EMSA has also made this the method of choice for following the subcellular distribution of DNA-binding proteins (3–5). This is particularly important when analyzing transcription factors, many of which are regulated by cytoplasm to nucleus translocation (reviewed in ref. 6).

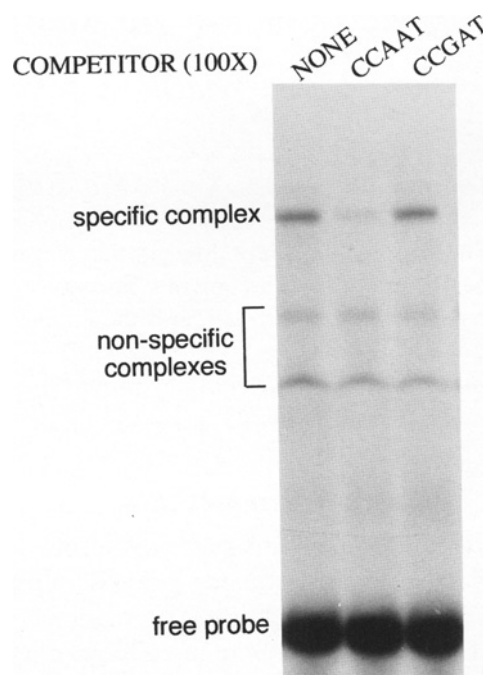


Fig. 1. Electrophoretic mobility shift assay of *Xenopus* CBTF. Extract from 0.5 stage 15 *Xenopus* embryos were incubated at 4°C with an end-labeled DNA probe corresponding to nucleotides –66 to –31 of the *Xenopus* GATA-2 promoter (3). Where added, the competitor was at 100-fold molar excess over the probe and both were added simultaneously. After 15 min the reactions were loaded onto a 4%, 0.25X TBE polyacrylamide gel. The free DNA and DNA–protein complexes were separated by electrophoresis at 200V for 2 h, the gel was dried, and the complexes visualized by autoradiography.

Although EMSA using crude cell extracts as a protein source provides very limited information about the nature of the interaction between the DNA and protein, it does provide a very useful assay for the amount of binding activity in any given sample. It is not possible, however, to be absolutely quantitative using this assay, when nonspecific competitors are included; these will have an affinity for the DNA-binding protein being assayed and necessarily cause errors in the estimation of activity present. Nevertheless, EMSA is an important semi-quantitative assay and is usually chosen to follow DNA-binding protein activity during purification (7–9).

Before embarking on the purification or detailed biochemical characterization of an unknown DNA-binding protein, it is prudent to undertake a preliminary analysis. In such a study, we attempt to discover how many subunits are in the DNA-binding complex and their size. Two basic strategies are used to identify the composition of such complexes from crude cell extracts, both relying on the bound DNA sequence to detect the proteins. A variety of techniques can be used to crosslink proteins covalently to the DNA probe. Alternatively, a blot of protein from a sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel may be incubated with a radiolabeled double-stranded oligonucleotide-binding site under conditions where the interaction is specific, and the resulting complex visualized by autoradiography.

2.2. DNA-Protein Photocrosslinking

Covalent crosslinks between protein and DNA have been produced by a variety of methods (10). Although a number of the techniques have relied on photocrosslinking, for example, ultraviolet (UV) laser irradiation of unmodified DNA and UV-irradiation of azido-modified DNA, the technique most commonly applied to analysis of a complex in a crude extract relies on substitution of 5-bromo-2'-deoxyuridine (BrdU) for thymidine in the oligonucleotide binding site and covalent crosslinking by UV irradiation. In our experience, when working with crude or partially purified preparations of protein, it is necessary to use preparative EMSA to obtain sufficiently pure DNA-protein complex for clear results. The BrdU-substituted probe may be produced in two ways. If the exact nucleotides involved in contact with the protein component are not known, then a mixture of substituted probes are produced by "filling in" a partial oligonucleotide duplex with Klenow fragment and a nucleotide triphosphates (NTP) mixture containing dTTP and BrdUTP (typically at a 3:1 molar ratio) and one or more radiolabeled NTPs. If, however, a T in the binding site is known to contact the protein, then an oligonucleotide may be synthesized with the BrdU substitution at this position. The resulting duplex

is then radiolabeled as usual. Either type of probe will work, but the "mixed probe" approach is more likely to reveal proteins in a complex not directly involved in contact with the DNA. However, this type of probe may contain more than one substitution and hence crosslink to two proteins, or more than once to a single polypeptide, causing anomalous migration in subsequent SDS-PAGE analysis.

The DNA-protein complexes are then formed on the substituted probe as usual and separated from free probe and nonspecific complexes by EMSA. The resulting gel is exposed to UV on a transilluminator or crosslinker to link the DNA and protein covalently. The gel is subsequently exposed wet for autoradiography, and the specific complex cut out of the gel and eluted. This DNA-protein complex is then precipitated prior to running on SDS-PAGE. After fixing, the gel is dried and the complex visualized by autoradiography.

The attachment of DNA to a protein may cause it to run anomalously on an SDS gel, although often it has little effect (*see Fig. 2*). With internally labeled probes, it is possible to remove much of the DNA by gentle treatment of the complex with DNase I prior to SDS-PAGE, but this often results in a much reduced signal. Although crosslinking can be criticized for sometimes giving an inaccurate estimate of the molecular weight of the protein(s) in a complex with DNA, it does have the advantage of identifying components of a complex that are not in immediate contact with DNA. This type of information is vital when designing purification strategies for DNA-protein complexes. A similar crosslinking approach has been used to determine which nucleotides in a binding site contact which protein in a multi-subunit transcription factor. In this case, as well as BrdU substitution, BrdC and BrdG were used to replace bases at known DNA-protein contact points (revealed by interference techniques). After crosslinking and SDS-PAGE of each singly substituted oligonucleotide, it was possible to identify which base substitutions resulted in crosslinking to which molecular weight species of the complex (11).

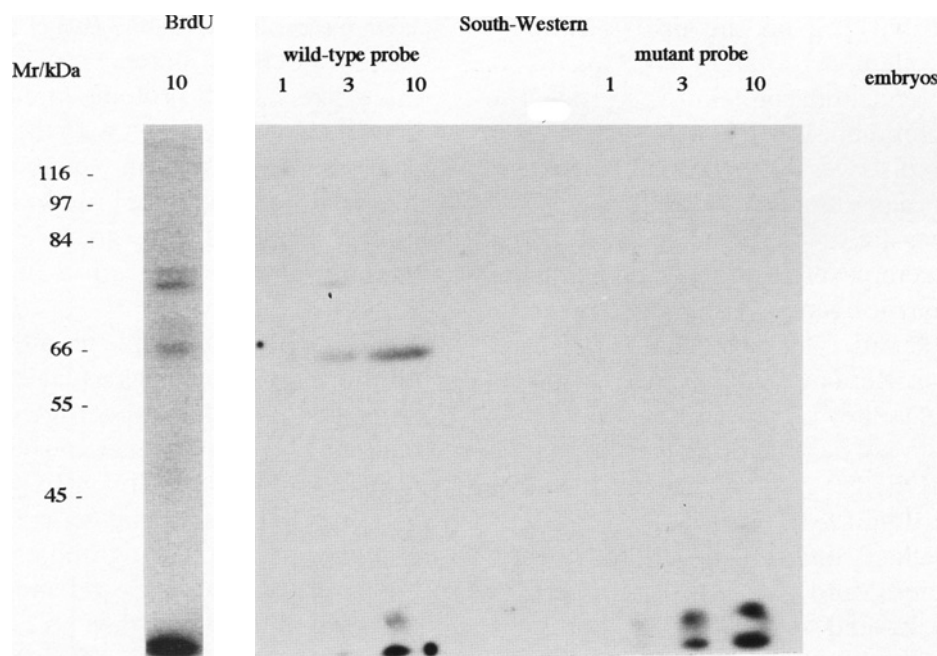


Fig. 2. **(Left)** Bromodeoxyuridine crosslinking of *Xenopus* CBTF. Extract from 10 oocytes was mixed with an internally labeled, BrdU-substituted DNA probe corresponding to –66 to –31 of the *Xenopus* GATA-2 promoter (3). Specific DNA–protein complexes were separated on a 4%, 0.25X TBE gel and irradiated with UV at 256 nm for 20 min. The complexes were visualized by autoradiography, cut from the gel, eluted into SDS-PAGE loading buffer and run on a 10% Laemmli gel. The gel was dried and the protein–DNA conjugate visualized after exposure on a storage phosphor screen. **(Right)** South-Western analysis of *Xenopus* CBTF. Extract from 1, 3, or 10 oocytes was run in duplicate on the same 10% Laemmli gel as the BrdU crosslinked products and transferred to nitrocellulose overnight in 10 mM CAPS, pH 9.8, 10% methanol. The duplicate filters so prepared were immediately blocked and probed. The probe was end-labeled DNA corresponding to nucleotides –66 to –31 of the GATA-2 promoter for the “wild-type” half of the filter and the same sequence but with the core altered from CCAAT to CCGAT for the “mutant” half. This point mutation inactivates the GATA-2 promoter in vivo and abolishes binding of CBTF in EMSA. Washing was for four by 15 min, this extended wash removed all nonspecific binding. The DNA–protein complexes were then visualized by autoradiography.

2.3. South-Western Blotting

A complementary method that gives accurate information about the molecular weight of the DNA-binding protein of a DNA–protein complex (but none on proteins not in direct contact with the DNA) is South-Western blotting (12). To obtain reliable results from the technique, it is necessary to have available as control a point-mutated version of the oligonucleotide probe that does not form the specific complex of interest. This is usually tested by EMSA prior to embarking on a South-Western investigation.

Crude cell protein is first separated by SDS-PAGE and then transferred electrophoretically to a

nitrocellulose membrane. During the transfer process, a fraction of the protein may re-fold sufficiently to bind its cognate DNA. If this does not occur, correct re-folding may be enhanced by treating the filter with Guanidinium-HCl to denature the protein and then re-folding in decreasing concentrations of this agent (13). Duplicate filters are prepared and nonspecific DNA-binding sites blocked by incubation with salmon-sperm DNA, Denhardt's reagent, and poly(dI.dC). The filters are then incubated with radioactively labeled oligonucleotide probes, either the bona fide binding site or the mutant. Filters are washed to remove unbound or weakly bound probe and exposed for autoradiography. Specific com-

plexes will show as differences between the bona fide and mutant binding sites.

South-Western blotting will fail if a heteromeric protein complex is necessary for DNA binding. Also, if multiple, specific bands are observed, then this technique shows that a family of proteins can recognize a particular DNA sequence. In the example shown, the complex is seen to include at least two proteins by BrdU crosslinking and the smaller of these is demonstrated by a South-Western blot to be the DNA-binding subunit of this transcription factor (**Fig. 2**).

2.4. Functional Analysis of Transcription Factors

Relating a DNA-protein complex visualized by EMSA or footprinting to a particular biological function is a difficult but very necessary undertaking when DNA-protein interactions in less well-characterized systems are being considered. If transcription factors are under study, it may be possible to make use of interference footprinting data to test whether a complex has function. By making point mutations in the DNA-binding site of a complex based on such data, it is possible to generate and test by EMSA a series of binding sites with various abilities of complex formation. The functional analysis of this series of binding sites is then carried out by transfection.

The mutations are either introduced into the promoter/enhancer under test or placed upstream of an heterologous promoter. In either case, these control a reporter gene. The constructs are transfected into the cell type under study and the effect of the mutations of the binding site on transcription assessed. If the ability of a DNA sequence to form complex and its transcription-regulation activity correlate, this is strong evidence that the observed complex is responsible for the activity seen. Such analysis showed that the ISGF3 complex was the most important of three formed on the interferon stimuable response element (ISRE) during stimulation of transcription of the 6-16 gene by type I interferons (**14**). Although this is strong evidence for a functional role of a particular protein, it is sometimes possible to obtain more direct evidence.

If purified transcription factor is available in sufficient quantity then supporting evidence to that of the aforementioned type can be gained from adding the factor to an in vitro transcription system and thus assaying its activity directly. A nuclear extract is prepared from cells either lacking the factor under test or from which the factor has been removed by DNA-affinity chromatography (**15,16**). A variety of methods for making transcriptionally active extracts may need to be tried for a particular DNA-protein complex. Into this extract, a DNA template containing the binding site of interest within an appropriate promoter and reporter context is added, together with NTPs. Transcription is then allowed to occur in the presence and absence of the putative transcription factor. Transcripts may be assayed in a number of ways. If a supercoiled template is used, then primer extension (**17**) or RNase mapping (**18**) of the resulting transcript may be used to test both the amount of transcript produced and that the in vivo start of transcription is used. Transcription initiation at cryptic start sites present in plasmids is prevalent in eukaryotic extracts.

Alternatively the "run-off" transcription assay may be employed. Here, the template is cut with a restriction enzyme at a point downstream (usually 100–250 bp) of the transcription start site. One or more of the nucleotide triphosphates added is radioactively labeled; the radioactive transcript synthesized is then of a discrete length as the polymerase "runs off" the end of the template. The amount of transcript and start of transcription may then be assayed directly by running the products on a sequencing-type gel. In practice, the tendency for transcription to start at cryptic initiation sites within plasmids means that both of these types of analysis give a high background of nonspecific signal. This may be overcome by placing a G-free cassette downstream of the promoter sequences under test. This is a stretch of DNA that contains no guanosine residues. The run-off transcription reaction is then carried out in the absence of guanosine triphosphate and, as all other transcripts will rapidly abort when transcribing a G, only those initiating at the correct site will be seen. This method of transcription factor assay is now widely used.

A final method for assaying the activity of a transcription factor relies on isolating the cDNA and the gene encoding it, and then inactivating the gene by insertional mutagenesis (19,20) either in cultured somatic cells, or embryonic stem cells from which mutant embryos lacking a functional transcription-factor gene may later be generated. Subsequent comparison of transcription levels of target genes of the transcription factor in normal and mutant cells will give information on the functional significance of that factor, but any effects observed may be indirect. If animals containing the inactivated gene are made, then phenotypic effects may be noted. Such a strategy, although long-term and involved, is often the only way of analyzing the function of, for example, a developmentally important transcription factor, (21).

In practice, a number of DNA-protein complexes may interact at one binding site (22), possibly controlling gene transcription in response to more than one intracellular signaling pathway. This complexity makes it extremely difficult to be sure, in some cases, of the exact contribution of any single DNA-protein complex to the final transcription rate of a gene and all three of the types of analysis previously outlined may be needed before any firm conclusions may be drawn.

3. Locating the DNA-Binding Site: DNA Footprinting

In order to confirm that a protein is indeed binding to a particular region of DNA and to discover the precise sequence to which it binds, footprinting techniques are most frequently used. The DNA region under investigation is uniquely end-labeled and allowed to interact with a source of protein, which may range from crude nuclear extract to a pure preparation. The DNA-protein complex is then treated with an enzymatic or chemical reagent that partially cleaves the DNA. The bound protein protects the region of the DNA with which it interacts from attack by the reagent. Denaturing electrophoresis and autoradiography of the degraded DNA (in parallel with a DNA sequencing ladder) identifies the protected region as a gap in the continuous bands of the digestion products (Fig. 3). If one is to gain maximum

advantage from the technique, then wherever possible the bands on the sequencing gel should be analyzed quantitatively (e.g., on a phosphor-imager) because densitometry of autoradiographs can give misleading results owing to nonlinearity of X-ray film.

3.1. Enzymatic Footprinting Methods

3.1.1. DNase I Footprinting

DNase I was the first reagent to be used in a DNA footprinting experiment (23). Because DNase I is large when compared to most other footprinting reagents, its attack on the DNA is more readily prevented by steric hindrance. Moreover, the reaction is carried out under mild conditions that are unlikely to perturb the complex. Thus, this reagent is the most likely of all those used for footprinting to detect a specific protein-DNA interaction. For example, the interaction of *Xenopus* upstream binding factor with the ribosomal DNA enhancer can only be detected at present using DNase I as a footprinting reagent (24). Similarly, DNase I footprinting can be used to detect a protein-DNA complex which is undetectable by EMSA and has been used as an alternative to that assay for analysis of proteins in crude cellular extracts.

DNase I does, however, have some disadvantages as a footprinting reagent. It has a diameter of roughly 40Å; its size, therefore, prevents it from cutting immediately adjacent to a bound protein. This means, of course, that the footprint observed will extend beyond the genuine contacts between the DNA and the protein, although it is possible to allow for this effect (25). In addition, DNase I footprinting generally requires an excess of DNA-binding protein over the DNA fragment, to ensure the DNA-binding site is saturated; this may limit the application of this technique where only small amounts of protein are available—for example, in developmental studies when only small numbers of embryos can be obtained. Finally, cleavage of DNA by this enzyme is variable; some sequences are very rapidly digested, while others (principally A or T tracts) are untouched even with prolonged incubation. Thus, an uneven ladder of digestion products is seen after electrophoresis, sometimes limiting the interpretation of results, because protection within a poorly cleaved region may not be readily apparent.

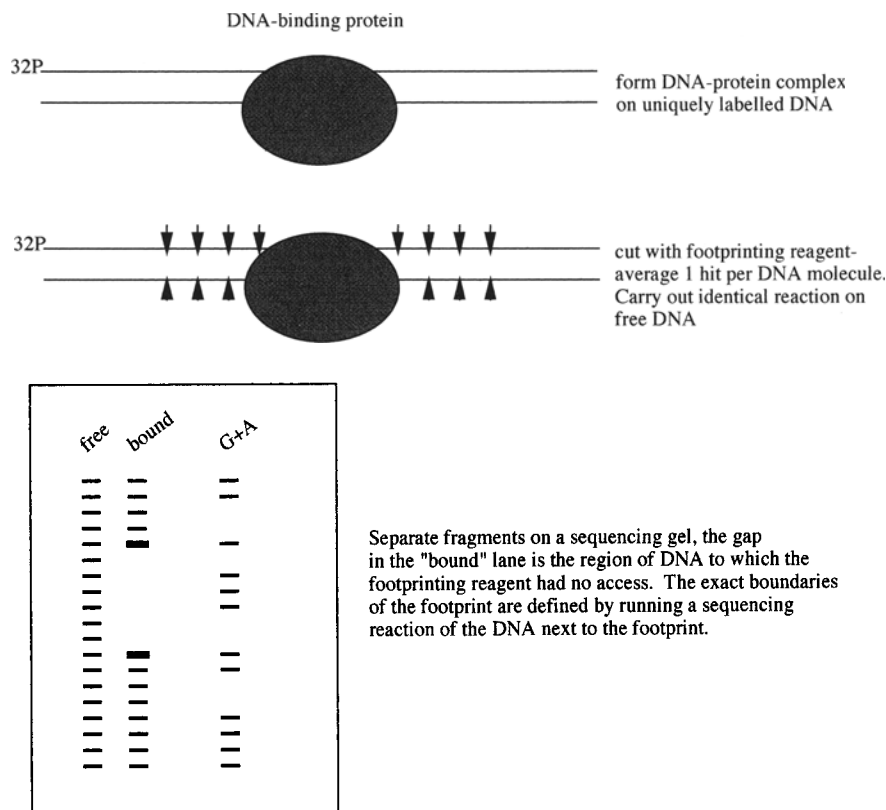


Fig. 3. Schematic diagram of a footprinting reaction.

3.1.2. Exonuclease III (*ExoIII*)

A second enzyme commonly used for footprinting is *ExoIII* which has 3'-5' exonuclease activity (26). After a protein-DNA complex has been formed, this enzyme is used to digest both strands processively. The protein blocks the action of *ExoIII* and a double-stranded DNA region protected by the protein remains. The labeled DNA is then run on a sequencing gel and the size of the remaining fragments corresponds to the ends of the protected region. Because all the free DNA is digested, there is no background ladder of digestion products observed, providing a distinct advantage over other footprinting techniques. Assigning the observed bands to the + or - DNA strands can be achieved either by labeling a unique end of the DNA fragment, or cleavage of the label from one end if a convenient restriction enzyme site is present. The processive mode of action of *ExoIII* allows

it to cut into a region of DNA with low affinity for its cognate-binding protein. This potential disadvantage can be overcome by carrying out a time-course of *ExoIII* digestion whenever a new interaction is being studied. Indeed, the ability of *ExoIII* footprinting to reveal (and distinguish) both strong and weak interactions can be an advantage. As with DNase I, *ExoIII* can be used to detect protein-DNA complexes that are not revealed by EMSA (27). Nevertheless, only interactions with a half-life longer than the application time of the *ExoIII* can be detected by this method (26,28). One further limitation of *ExoIII* footprinting is that it is only able to reveal the protected region at the 3' end of each strand (which need not be equivalent to the 5' boundary on the opposite strand, although it is often assumed to be so). However, for preliminary (low-resolution) studies, this is often sufficient.

3.2. Chemical Footprinting Methods

Chemical footprinting methods all have one clear advantage over enzymatic ones: The small size of the reagents used means that they can cut the DNA very closely indeed to the bound protein, resulting in higher resolution than can be achieved with DNase I or *ExoIII*. The possibility should be borne in mind, however, that chemical probes, unlike enzyme probes, may chemically modify the protein component as well as the DNA component of a complex. The hydroxyl radical is one chemical reagent commonly used, and is particularly useful for high resolution studies because it is comparable in size to a water molecule. The exact mechanism whereby these radicals cleave DNA is unclear, but is thought to involve the abstraction of a hydrogen atom from a sugar in the DNA backbone and the breakage of the strands by secondary reactions of the sugar radical (29). The principle involved is exactly the same as DNase I footprinting discussed earlier. However, unlike DNase I, hydroxyl radicals have almost no sequence preferences for cutting and will work in a wide variety of conditions. One disadvantage of using hydroxyl radicals as a footprinting reagent is when purified protein samples have been stored in glycerol; this is a radical scavenger and inhibits cutting when present at concentrations higher than 0.5%. Other chemical-footprinting reagents include dimethyl sulfate (DMS), a reagent that is fairly specific for purine bases, modifying the N7 of guanine and the N3 of adenine (30).

One problem that is often encountered when attempting to footprint a DNA binding protein is that of low occupancy, i.e., if not all the DNA molecules in the sample have a protein bound, then there will be a large background owing to free DNA, which will reduce the clarity of the footprint. This will often be the case when using crude extracts, but problems can also occur with pure proteins, especially if the complexes are not very stable. One possible solution is to ensure that there is a large excess of the protein compared to the DNA, but this can itself give rise to problems of multiple binding, where the protein binds to additional sites on the DNA in a relatively non-

specific manner. (This becomes more likely as the size of the DNA fragment is increased.) Another approach is to purify the free- and bound-DNA fractions following an EMSA on a nondenaturing gel, before carrying out the footprinting reaction; this does at least ensure that the protein and DNA are present in stoichiometric amounts, and may be particularly useful if working with crude extracts of protein, or if the concentrations are unknown. However, if the complexes are weak or unstable, there is no guarantee that the protein will remain fully bound to the DNA following extraction from the gel.

A novel variation on this theme is to perform the footprinting reaction *in situ* in the nondenaturing gel. This approach has been used very successfully with the footprinting reagent copper phenanthroline (31) and has found application in studying a wide variety of DNA-protein interactions (*see, e.g., ref. 32*). A major advantage of this variation on the standard footprinting procedure is that kinetically weak complexes should be maintained in the gel. Additionally, if multiple complexes are formed, DNA from each can be separately extracted from the gel after footprinting and the protected sites in each complex can be investigated.

3.3. DNA "Footprinting" as a Probe of Structure

Traditionally, DNA footprinting has been considered to be a relatively straightforward procedure for mapping the protected sites on the DNA, as judged by weak or absent bands in a DNA sequencing ladder where the protein is bound. However, proteins can also distort the structure of DNA when they bind, sometimes quite severely, and in such cases, what might be seen is an increase in the intensity of certain bands, rather than a decrease.

A very clear example of this phenomenon is provided by the hydroxyl radical footprinting of the type I DNA methyltransferase *M.EcoR124I* (33). In this case, binding of the protein to a DNA fragment containing its recognition sequence yields a clear and extensive footprint, amounting to over two turns of the double helix on both strands. Within this footprint, however, is found a region of 2–3 nt that is very highly modified,

indeed at considerably higher levels than are found in the absence of protein (**Fig. 4**). These sites correspond to the bases that are modified by the methyltransferase, and it was concluded that the DNA structure here is severely distorted, most probably arising from “flipping out” of the two target bases in a similar way to that seen in the crystal structure of a number of DNA-modifying enzymes (34). Similarly, enhancements in the hydroxyl radical footprint have been taken as evidence for structural distortion of DNA in a type 1 transpososome (35).

Markedly increased intensity of bands in the DNA ladder are often seen with DNase I footprinting, particularly at the 3' boundary of the footprint. Even though the bands are much more intense than in the free DNA track, this is unlikely to be owing to a conformational change in the DNA because the effect is very common, and typically occurs at the edge of the footprint, i.e., outside of the binding site rather than within it. It most likely represents a favorable interaction between the DNase I and the bound protein, which would enhance the cutting frequency at that site. Nevertheless, such an effect is quite helpful in visualizing the precise boundaries of the footprint.

3.4. Footprinting In Vivo

Although the footprinting techniques previously discussed can identify precisely the binding site for protein on a DNA duplex, they all suffer from the drawback that the complex is formed in vitro. As a result, cofactors may be missing or the precise ionic conditions found within the cell may not be duplicated and thus the interaction is in some way artifactual. This criticism can be overcome by carrying out DMS modification of DNA in vivo, because DMS can penetrate plasma and nuclear membranes (36).

In essence, cells are treated for a short period with DMS and genomic DNA isolated. The modified guanine residues are then cleaved with piperidine. In order to increase the amount of the specific sequences of interest, ligation-mediated polymerase chain reaction (PCR) can be used to amplify these with appropriate primers (37,38). The cleaved DNA is then separated on a sequenc-

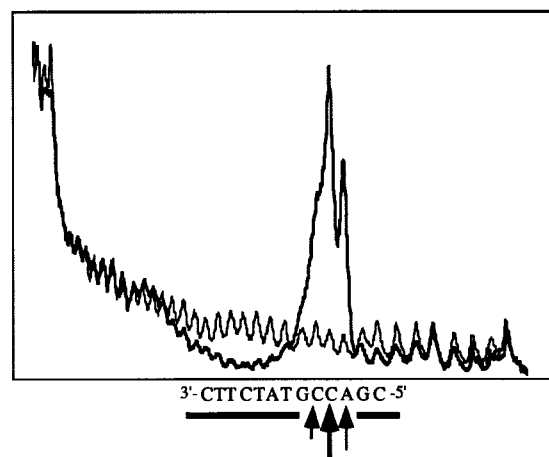
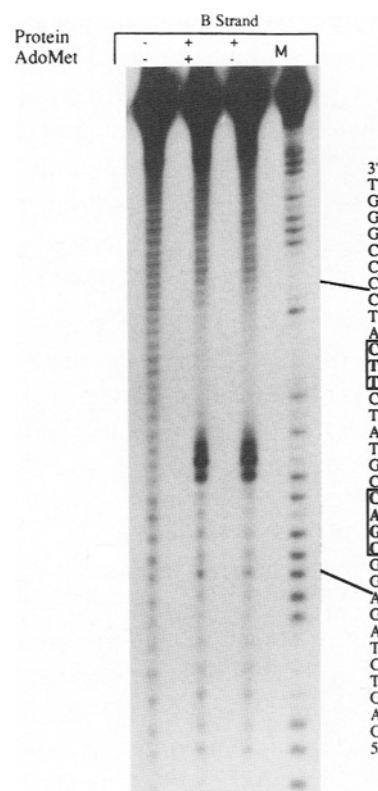


Fig. 4. Hydroxyl radical footprinting. The reaction shown is for the complex of the DNA methyltransferase *M.EcoR124I* and a 57 bp DNA duplex containing the recognition sequence, with one strand labeled. The extent of cleavage was obtained by scanning the gel on a phosphorimager, in order to quantify the level of protection and enhancement at different sites.

ing-type gel and transferred electrophoretically to a nylon filter. The particular sequences of interest are then visualized by hybridization of the filter to a

radioactively labeled single-stranded DNA probe and subsequent autoradiography. Using this method, it has been possible to identify footprinted regions of promoters that are not seen *in vitro* (36), and this method has proven particularly useful in studying allele-specific transcriptional activity (39).

4. Identification of Specific Bases Involved in Protein Binding

The DNA footprinting procedures previously referred to give information on the region of the DNA that is protected when the protein binds. An alternative (and complementary) approach is to modify the bases of the DNA and assess which base modifications prevent (or reduce) binding. Such methods give a direct measure of the contribution of individual nucleotides to the binding affinity, and can thus identify specific molecular interactions in the complex. Moreover, because often it is only specific groups of the nucleotide that are modified in the DNA (for example, the phosphate group, the sugar, or a specific group on the base), then information is provided on the precise nature of the intermolecular contacts.

4.1. Interference Footprinting

The most common experiments of this kind are represented by the “interference” footprinting techniques, in which the DNA sequence is chemically modified at a low level (typically yielding approximately a single modified nucleotide per DNA fragment). The protein is then bound to the modified DNA, and the mixture subjected to nondenaturing gel electrophoresis to separate free and bound DNA, both of which are then subjected to a strand-cleavage reaction (in which the DNA is cleaved specifically at the sites of modification) and subsequently applied to a sequencing gel. By determining the sites of modification in the free and bound DNA, the contribution of each nucleotide to the binding affinity for the protein will be established. Those nucleotides modified at crucial sites will be under-represented (or absent) in the DNA sequences from the bound fraction, whereas those nucleotides whose modification does not affect protein–DNA complex formation will be well represented in the bound DNA sequences.

The “interference” method may be thought of as the reverse of “protection” footprinting, the

crucial difference being that in the latter case, protein binding is carried out before chemical modification of the DNA, whereas in the former, the DNA is modified first and then “challenged” for protein binding. As in the “protection” footprinting experiments described earlier, the DNA strands must be separately end-labeled with ^{32}P ; chemical modification, DNA binding, and subsequent DNA sequencing should be carried out in parallel reactions with each strand separately labeled, together with a Maxam and Gilbert sequencing reaction on the DNA fragment to enable identification of the sites of modification. Again, if possible, the bands on the sequencing gel should be analyzed on a phosphorimager to obtain reliable data on the relative population of modified DNA molecules in the “free” and “bound” fractions.

The principle of the interference technique is similar, regardless of the reagent used, although the experimental details differ, as does the interpretation of the results. Hydroxyl radical modification, for instance, itself leads to cleavage of the DNA, and thus no separate step is required prior to running the sequencing gel. Most other commonly used reagents (e.g., DMS or ethyl nitrosourea [ENU]), on the other hand, require an additional step to convert the nucleotide modification into a strand cleavage so that the position of the modification can be determined on a sequencing gel (30,40).

Hydroxyl radical modification leads to cleavage via the removal of the entire nucleoside, and thus eliminates all interactions of the protein with the missing base and/or sugar (41,42). Because the reagent is relatively insensitive to base sequence, it has the advantage of being able to probe all nucleotides in a DNA-binding site. However, the missing nucleoside can also change the conformation of the DNA or enhance its flexibility, which can complicate the interpretation of a hydroxyl radical interference experiment. For example, the perturbed DNA structure may actually enhance binding in some cases, as seen in the case of RNA polymerase (43).

More localized contacts of a DNA-binding protein with specific groups on the bases can be probed by interference experiments using the reagent DMS. As stated earlier, DMS methylates

the N7 position of guanine and the N3 of adenine and thus is able to probe interactions in the major and minor grooves of DNA. This does, however, mean that DMS is unable to provide any information on the interaction of a DNA-binding protein with pyrimidine bases in a DNA-binding site. Nevertheless, methylation interference has proved a very useful technique in studying numerous DNA-binding proteins; for example, in the case of the *c-fos* serum-response element binding factors, methylation interference experiments identify a subset of the bases suggested by methylation protection as being essential for binding (44).

Although not generally leading to sequence specificity, the interaction of the phosphate groups on the DNA backbone with a DNA binding protein are often critical in maintaining the high affinity of binding. ENU is widely used to probe such interactions, and can identify critical phosphate groups that are important for DNA binding. Ethylation interference experiments have been used to probe the interaction of the *Escherichia coli* MetJ repressor and the phage-434 repressor with their respective operator sequences (45,46). In both cases, they identified phosphate groups on the DNA that correlated with interactions that could be deduced from the crystal structure of the DNA-protein complexes.

4.2. Base Analogs

The interference methods previously described can be carried out using either a natural (cloned) DNA-binding site, or a synthetic-oligonucleotide duplex bearing the appropriate sequence. It is worth noting that an alternative way to probe the effect of individual base modifications on the DNA-binding affinity of a protein is to incorporate modified bases at specific sites during the synthesis of the two strands that are used to form the duplex. This, naturally, is more time-consuming (and costly!) than the chemical modification procedures previously described, because individual DNA duplexes have to be constructed for each modification to be studied. Moreover, each modified DNA duplex has to be tested individually for DNA binding, usually by a gel-retardation assay (EMSA).

Thus, this type of analysis is generally most useful when a DNA-binding protein has already been characterized in some detail, and the potential base changes to be analyzed are clearly indicated by earlier experiments. The results are generally more clear cut than those obtained by interference "footprinting," and can yield quite accurate values for the relative affinity of the protein for DNA sequences lacking specific groups on a base (or even the entire base), from which the contribution of the relevant interaction(s) to the binding energy can be deduced (47). There are now a large number of base analogs that can be used in this way in which functional groups are deleted (e.g., inosine, purine, 2-amino purine) or modified (e.g., 6-thio G, and various methylated bases such as 5-methyl C, N6-methyl G), as well as the more traditional approach of simply substituting one base by another of the standard bases A, G, C, or T (or indeed, U). If uracil has been incorporated into the sequence, then treatment of the oligonucleotide with the enzyme uracil-N-glycosylase provides a convenient way to remove the entire base from a given site.

4.3. Binding-Site Selection

At the other extreme, it is possible to synthesize an oligonucleotide that is degenerate at a number of sites, i.e., one which contains a random mixture of the four standard bases at certain positions in a given sequence. By selecting those sequences that are preferentially bound by the protein, typically using EMSA, the preferred bases at selected sites in the sequence can be established. Many variations of the basic technique can be found, known by various names such as binding site selection analysis (BSSA) or sequence selection by exponential enrichment (SELEX). Using low-degeneracy oligonucleotide mixtures, and direct chemical sequencing of bound and free DNA following EMSA, Gogos et al. (48) were able to identify preferred bases in the binding site of the CF2 zinc-finger protein from *Drosophila*. In this method, the use of solid-phase sequencing considerably simplifies the procedure (49). Other strategies can be adopted that make use of PCR or DNA cloning after the initial selection step, using a highly degenerate

oligonucleotide mixture initially; multiple rounds of selection can thus be used to amplify small sequence preferences to determine the optimum DNA-binding site (50,51).

5. Quantitative Analysis of Binding

When working with purified and well-characterized DNA-binding proteins, one would usually like to know the equilibrium binding constant that characterizes the interaction of the protein with one or more DNA sequences. Many techniques are available for such studies, the most commonly employed being gel-retardation assays (EMSA) which, as discussed in **Subheading 2.1.**, can also be used more qualitatively to characterize interactions in protein extracts or to aid in the separation of free and bound protein when using many of the techniques referred to already. Spectroscopic methods (principally fluorescence and circular dichroism), and quite unrelated techniques such as surface plasmon resonance, each have advantages for the analysis of binding.

Whichever method is used, it is vital to know the protein concentration fairly accurately. Clearly, the protein must be pure according to SDS-PAGE; it should also have a good UV spectrum with no sign of contaminating nucleic acid. It is surprising how often this is overlooked; not only would contaminating DNA affect the observed binding to the DNA-binding site, but it would also give rise to errors in determination of protein concentration. Assuming the sample is spectroscopically pure, a reasonable estimate of the protein concentration will be obtained if the molar-extinction coefficient of the protein is calculated simply by summing the product of the number and extinction coefficient of each aromatic amino acid at a given wavelength, normally 280 nm.

For many purposes, it is often necessary to know the accurate concentration of the DNA. For defined sequences, this can be done by adding the contribution from each of the bases present in the sequence. However, hyperchromicity should be taken into account and it is advisable to carry out snake-venom phosphodiesterase treatment of the DNA to allow correction for this factor (52) because the absorbance of the DNA will be lower

than the sum of its component bases, even if the DNA duplex is "melted" at high temperature (when some residual hyperchromicity will always remain owing to base stacking).

5.1. EMSA of Purified Proteins

EMSA are widely used to quantify the affinity of a protein for a given DNA sequence. Two rather different approaches can be used, both relying on the titration of a protein of known concentration into a fixed concentration of DNA. Often, the DNA concentration is very low (compared with the K_d for the interaction) and is not accurately known. In this case, relatively high concentrations of protein will be required to achieve significant binding, and the free protein concentration will be close to that of the total protein concentration (53). If so, the analysis is very similar to the Michaelis-Menton analysis for enzyme-substrate reactions (in which it is assumed that the enzyme concentration is negligible compared with the substrate concentration). Similar procedures for fitting the data can be used, and likewise an estimate of the K_d can be obtained from the concentration of protein needed to shift 50% of the DNA (cf. the relationship between K_m and V_{max}).

One potential drawback of this approach, however, is that a 1:1 stoichiometry is assumed (together with the assumption that all the protein is fully active). In order to check this, it is advisable to carry out the gel mobility assay at a higher concentration of DNA (and consequently, of protein). At sufficiently high concentrations ($[DNA] \gg K_d$), binding will be linear with protein concentration until the stoichiometric point is reached in the titration, at which point all the DNA will be bound (**Fig. 5**). If both the DNA and protein concentrations are known, the ratio of the two concentrations at this point gives the stoichiometry, n . Apparent values of $n > 1$ could indicate that more than one protein is binding to the DNA; an alternative explanation is that not all of the protein is active. (One way to attempt to distinguish these possibilities is to repeat the experiment with different preparations of the protein.) It is often apparent when more than

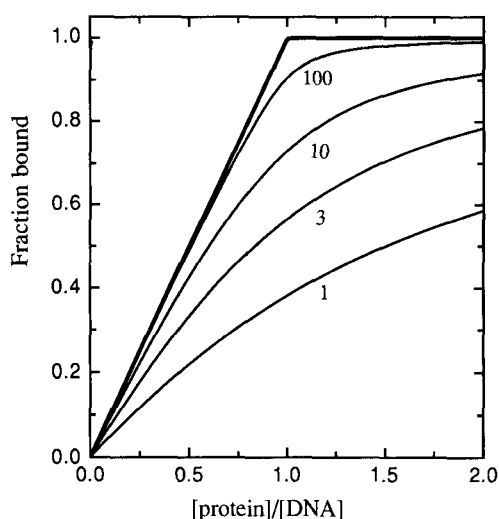


Fig. 5. Theoretical curves for protein binding to a single site on DNA for different (fixed) DNA concentrations. The binding curves for DNA concentrations equal to 1, 3, 10, and 100 times the K_d are shown. The curve is also shown for infinitely tight binding (heavy line).

one protein is binding to the DNA as multiple bands may be seen corresponding to the various complexes that can form (53). However, the lack of such bands does not necessarily indicate a 1:1 stoichiometry; for example, a single band is observed for highly cooperative interactions such as the multiple binding of gene 5 protein to oligonucleotides (54).

At intermediate DNA concentrations, only a fraction (f) of the DNA will be bound at protein concentration (P_0) corresponding to the stoichiometric point (Fig. 5). In this case, it can be shown that $K_d = P_0(1 - f)^2 / f$ and the equilibrium constant can be obtained without any approximations (55). This method does require the concentration of the DNA to be known, and for this reason, is often done with synthetic oligonucleotides, which are prepared in sufficient quantities for spectroscopic determination of concentration after labeling. For the former approach, small quantities of a cloned DNA fragment will suffice.

It is also possible to measure the kinetics of dissociation of a protein-DNA complex by EMSA. To do this, one can form a 1:1 complex, and compete off the protein from the labeled DNA by incubation with an excess of unlabeled DNA for

varying times. The amount of labeled complex should decrease exponentially with time, and a log plot will yield the dissociation rate constant, K_d (as long as the latter is significantly longer than the "dead" time of the experiment).

Competition assays can be used quantitatively to determine the relative binding affinity of a protein for two or more DNA sequences, for example, to compare binding to DNA sequences containing base analogs or point mutations. A stoichiometric (1:1) complex is formed with the canonical DNA sequence at a fixed concentration (C_1). From the concentration (C_2) of unlabeled competitor needed to dissociate 50% of the complex, the ratio of binding affinities for the two sequences can be obtained simply from the formula $K_1/K_2 = [2(C_2/C_1) - 1]$ (see ref. 52).

5.2. Surface Plasmon Resonance

Surface plasmon resonance (SPR) or related techniques such as resonant mirror devices can be employed to study the kinetics of protein-DNA interactions by attaching a DNA probe to the surface of a chip. Usually, the oligonucleotide is biotinylated so that it can be linked to streptavidin on the chip surface. The protein is allowed to flow over the surface of the chip, and the increase in mass due to protein binding measured (usually in arbitrary resonance units, or [RU]) as a function of time. Once bound, the protein can be dissociated from the DNA by flowing buffer over the chip, and the signal will decrease. Thus analysis of the time dependence of these two processes can, in principle, allow the kinetics of the interaction to be investigated, giving the rate constants for association and dissociation (k_{on} and k_{off}). Figure 6 shows one such example.

For complex binding (e.g., if a number of proteins bind cooperatively), analysis is not at all straightforward. Even for simple equilibria, however, where numbers are easy to obtain by fitting the curves to appropriate equations, one must exercise great caution in interpreting the "apparent" values obtained, which can be subject to errors of many different kinds. A number of self-consistency checks are available that should always be used before taking the derived parameters too seriously (56).

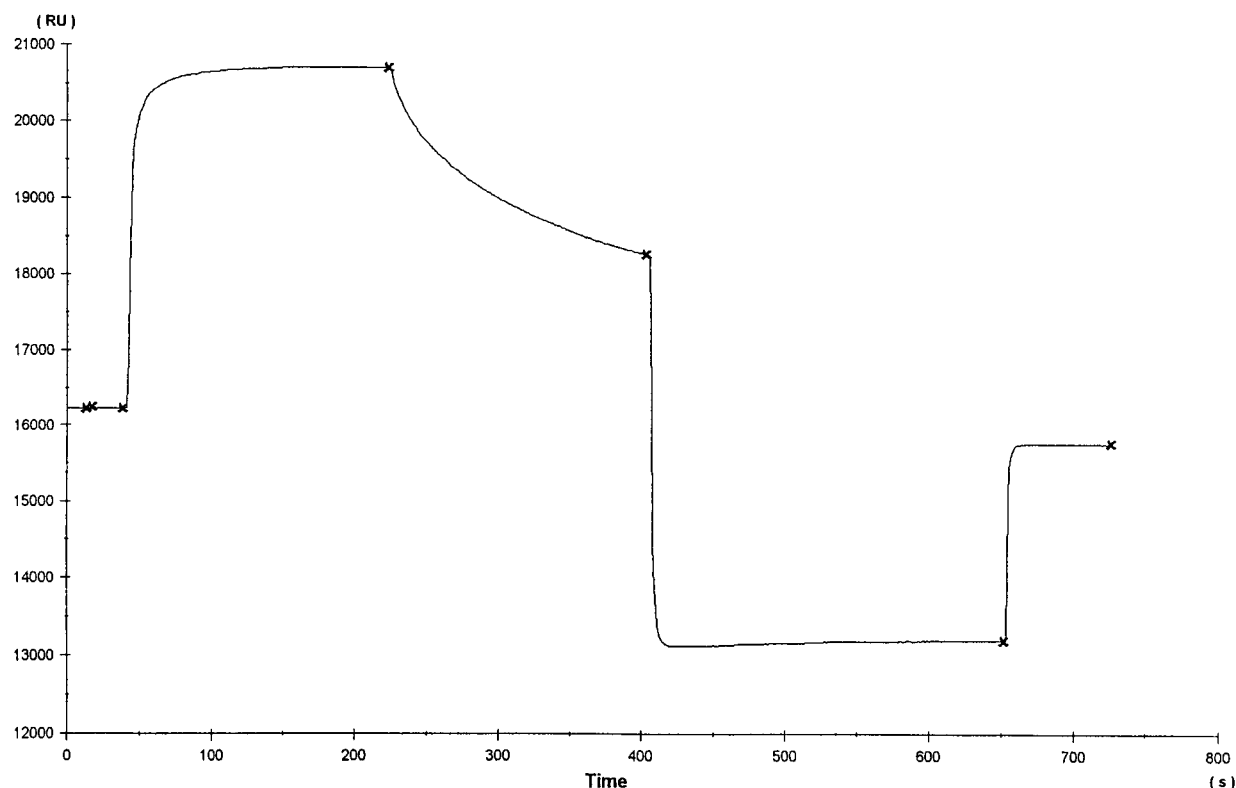


Fig. 6. Analysis of the binding kinetics of the DNA methyltransferase *M.EcoR124* by SPR. The first two phases of the sensorgram show binding of the protein to a 30-bp DNA sequence containing the recognition site, followed by dissociation of the protein from the DNA. Subsequent experiments in which the concentration of protein is varied allow the kinetic constants of the interaction to be obtained. In this example, the curves could be fitted with the values $k_{on} = 2.5 \times 10^5/s/M$ and $k_{off} = 2.6 \times 10^{-3}/s$ in good agreement with the measured equilibrium binding constant of $10^8/M$.

Nevertheless, SPR provides a very useful technique for comparative analysis of DNA binding proteins (e.g., using different DNA-binding protein mutants), and there are many ways in which the technique can be adapted. For example, the DNA-binding protein itself can be immobilized on the surface of the chip, and a variety of DNA sequences can be tested for their affinity to the protein. Although there are a number of techniques available for obtaining equilibrium binding constants, or dissociation-rate constants, SPR provides one of the few methods for the determination of association-rate constants directly.

5.3. Spectroscopic Methods

Virtually all the biophysical techniques that can be applied to proteins can also be applied very effectively to analyze protein–DNA interactions.

Among the most valuable are fluorescence and circular dichroism, because they are relatively straightforward. Nuclear magnetic resonance and X-ray crystallography, of course, are able to provide far more detail if circumstances are favorable, but are much more time consuming and dependent on very expensive instrumentation.

5.3.1. Fluorescence Spectroscopy

Fluorescence can be used in many different ways to study the interaction of a protein with DNA. Most simply, one looks for a quenching or a wavelength shift in the intrinsic protein fluorescence as DNA is titrated in (55), which might arise if a tyrosine or tryptophan were involved in the interaction. The resulting binding curve can be analyzed by the methods described in **Subheading 5.1**. Alternatively, a fluorescent probe

can be attached (for example, to a cysteine at a specific site in the protein). This has the advantages of greater sensitivity (assuming one sees any change at all), but one should use some other methods to check that the probe has not perturbed DNA binding.

An alternative is to use a fluorescence-competition assay; for example, the fluorescent probe ANS binds at the DNA-binding site of the methyltransferase *M.EcoR124I*, and competition with the specific DNA sequence produces a large decrease in signal as the probe is displaced into free solution (57). Fluorescence polarization provides another approach; the increase in molecular weight of the complex, compared to the free protein, will often lead to a decrease in the extent of depolarization of the emitted light, especially in the case of large complexes (58).

5.3.2. Circular Dichroism (CD) Spectroscopy

CD is less sensitive than fluorescence and thus is not particularly useful for deriving binding constants for DNA-protein interactions. It does, however, have a number of important uses in analyzing the structure of DNA-protein complexes (59). The far UV CD (170–250 nm) is sensitive to changes in the secondary structure of proteins and thus any structural changes due to DNA binding can be studied. In the near UV CD (250–300 nm), the signal derives principally from the DNA component, and changes here reveal conformational changes in the DNA when the protein binds. For example, the gene 5 protein produces a large decrease in the peak at 276 nm on binding to single-stranded DNA (58), suggesting a loss of base-stacking interactions owing to intercalation of aromatic side chains of the protein. In contrast, the methyltransferase *M.EcoR124I* shows an increase of over 50% in the CD signal of the DNA (60), which may be related to the large distortion in the DNA-binding site suggested by hydroxyl-radical footprinting of this complex (33).

6. Conclusion

We have aimed to provide a brief overview of a wide variety of techniques that can be usefully employed to study DNA-protein interactions,

from the identification of novel DNA binding proteins in crude cell extracts, to the detailed analysis of the purified proteins and their mode of interaction with specific DNA sequences. Of necessity, we have had to be selective in the methods chosen for inclusion. For the sake of illustration, many of the examples described are taken from studies being undertaken in our own work, but the literature in this area is vast.

Many topics have been left out owing to restrictions of space; for example, many of the traditional techniques of protein chemistry and protein structure analysis are all widely used in the study of DNA-protein interactions; e.g., chemical modification and limited proteolysis (61,62). Overexpression of DNA-binding proteins and their domains, though not covered here, are also increasingly important tools. For details of these and other relevant techniques, the reader is referred to ref. 63.

Acknowledgments

We are grateful to the Wellcome Trust and the Biotechnology and Biological Sciences Research Council (BBSRC), UK, for continued financial support for ongoing research in our two laboratories, to Iain Manfield for helpful comments on the manuscript, and to our many colleagues who have contributed to this work.

References

1. Garner, M. M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *E.coli* lactose operon. *Nucleic Acids Res.* **9**, 3047–3060.
2. Fried, M. G. and Crothers, D. M. (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**, 6505–6525.
3. Brewer, A. C., Guille, M. J., Fear, D. J., Partington G. A., and Patient, R.K. (1995) Nuclear translocation of a maternal CCAAT factor at the start of gastrulation activates *Xenopus* GATA-2 transcription. *EMBO J.* **14**, 757–766.
4. Bauerle, P. A. and Baltimore, D. (1988) Activation of DNA-binding activity in an apparently cytoplasmic precursor of NF- κ B transcription factor. *Cell* **53**, 211–217.
5. Dale, T. C., Imam, A. M. A., Kerr, I. M., and Stark, G. R. (1989) Rapid activation by IFN- α of a latent

- DNA-binding protein present in the cytoplasm of untreated cells. *Proc. Natl. Acad. Sci. USA* **86**, 1203–1207.
6. Whiteside, S. T. and Goodbourn, S. (1993) Signal transduction and nuclear targeting: regulation of transcription factor activity by subcellular localisation. *J. Cell. Sci.* **104**, 949–955.
 7. Treisman, R. (1987) Identification and purification of a polypeptide that binds to the c-fos serum response element. *EMBO J.* **6**, 2711–2717.
 8. Hofmann, J. F. X., Laroche, T., Brand, A. H., and Gasser, S. M. (1989) RAP-1 factor is necessary for DNA loop formation in vitro at the silent mating type locus HML. *Cell* **57**, 725–737.
 9. Perkins, N. D., Nicholas, R. H., Plumb M. A., and Goodwin G. H. (1989) The purification of an erythroid protein which binds to enhancer and promoter elements of haemoglobin genes. *Nucleic Acids Res.* **17**, 1299–1314.
 10. Welsh, J. and Cantor, C. R. (1984) Protein-DNA crosslinking. *TIBS* **9**, 505–507.
 11. Qureshi, S. A., Salditt-Georgieff, M., and Darnell, J. E., Jr. (1995) Tyrosine-phosphorylated Stat1 and Stat2 plus a 48 kDa protein all contact DNA in forming interferon-stimulated gene factor 3. *Proc. Natl. Acad. Sci. USA* **92**, 3829–3833.
 12. Singh, H., LeBowitz, J. H., Baldwin, A. S., and Sharp, P. A. (1988) Molecular cloning of an enhancer binding protein: isolation by screening an expression library with a recognition site DNA. *Cell* **52**, 415–423.
 13. Celenza, J. L. and Carlson, M. (1986). A yeast gene that is essential for release from glucose repression encodes a protein kinase. *Science* **233**, 1175–1180.
 14. Dale, T. C., Rosen, J. M., Guille, M. J., Lewin, A. R., Porter, A. G. C., Kerr, I. M., and Stark, G. R. (1989) Overlapping sites for constitutive and induced DNA-binding factors involved in interferon-stimulated transcription. *EMBO J.* **8**, 831–839.
 15. Jones, K., Kadonaga, J., Rosenfeld, P., Kelly, T., and Tjian, R. (1987) A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* **48**, 79–89.
 16. Cortesy, B., Claret, F. -X., and Wahli, W. (1990) Estrogen receptor level determines sex-specific in vitro transcription from the vitellogenin promoter. *Proc. Natl. Acad. Sci. USA* **87**, 7878–7882.
 17. Leonard, M. W. and Patient, R. K. (1991) Primer extension analysis of mRNA isolated from transfected cell lines, in *Methods in Molecular Biology*, Vol. 7; *Gene Transfer and Expression Protocols* (Murray, E. J., ed.), Humana, Totowa, NJ, pp. 297–306.
 18. Sambrook, J., Fritsch, F. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 19. Nizielski, S. E., Lechner, P. S., Croniger, C. M., Wang, N. D., Darlington, G. J., and Hanson R. W. (1996) Animal-models for studying the genetic-basis of metabolic-regulation. *J. Nutr.* **126**, 2697–2708.
 20. Rijkers, T., Peetz, A., and Rutler, U. (1994) Insertional mutagenesis in transgenic mice. *Transgen. Res.* **3**, 203–215.
 21. Tsai, F.-Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W., Strauss, E., and Orkin S. H. (1994) An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371**, 221–226.
 22. Thanos, D. and Maniatis, T. (1995) Virus induction of human IFN- β gene-expression requires the assembly of an enhanceosome. *Cell* **83**, 1091–1100.
 23. Schmitz, A. and Galas, D. J. (1978) DNaseI footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5**, 3157–3170.
 24. Le Blanc, C. B. and Moss, T. (1994) DNaseI footprinting, in *Methods in Molecular Biology*, Vol. 30; *DNA-Protein Interactions*. (Kneale, G. G., ed.), Humana, Totowa, NJ, pp. 1–11.
 25. Fairall, L. M. and Rhodes, D. (1992) A new approach to the analysis of DNaseI footprinting data and its application to the TFIID/5S DNA complex. *Nucleic Acids Res.* **20**, 4727–4731.
 26. Wu, C. (1985) An exonuclease protection assay reveals heat-shock element and TATA box binding proteins in crude nuclear extracts. *Nature* **317**, 84–87.
 27. Decker, T., Lew, D. J., Mirkovitch, J., and Darnell, J. E., Jr. (1991) Cytoplasmic activation of GAF, an IFN- γ regulated DNA-binding factor. *EMBO J.* **10**, 927–932.
 28. Loh, T. P., Sievert L. L., and Scott, R. W. (1990) Evidence for a stem cell-specific repressor of moloney leukaemia virus expression in embryonic carcinoma cells. *Mol. Cell. Biol.* **10**, 4045–4057.
 29. Shafer, G. E., Price M. A., and Tullius Th. D. (1989) Use of hydroxyl radical and gel electrophoresis to study DNA structure. *Electrophoresis* **10**, 397–404.
 30. Shaw, P. E and Stewart, A. F. (1994) Identification of Protein-DNA contacts with Dimethyl Sulphate, in *Methods in Molecular Biology*, Vol. 30; *DNA-Protein Interactions* (Kneale, G. G., ed.), Humana, Totowa, NJ, pp. 79–95.
 31. Papavasillou, A. G. (1994) 1,10 phenanthroline-copper ion nuclease footprinting of DNA-protein complexes in situ following mobility shift electrophoresis assays, in *Methods in Molecular Biology*, Vol. 30; *DNA-Protein Interactions*. (Kneale, G. G., ed.), Humana, Totowa, NJ, pp. 43–78.
 32. Flanagan, W. M., Papavassiliou, A. G., Rice, M., Hecht, L. B., Silverstein, S., and Wagner, E. K. (1991) Analysis of the Herpes Simplex Virus type 1 promoter controlling the expression of U_L38, a true late gene involved in capsid assembly. *J. Virol.* **65**, 769–786.
 33. Mernagh, D. R. and Kneale, G. G. (1996) High resolution footprinting of a type I methyltransferase

- reveals a large structural distortion within the DNA recognition site. *Nucleic Acids Res.* **24**, 4853–4858.
34. Cheng, X., Kumar, S., Posfai, J., Plugraht, J. W., and Richards, R. J. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357–369.
 35. Lavoie, B. D., Chan, B. S., Allison, R. G., and Chaconas, G. (1991) Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu type I transpososome. *EMBO J.* **10**, 3051–3059.
 36. Becker, P. B., Ruppert, S., and Schutz, G. (1987) Genomic footprinting reveals cell type-specific binding of ubiquitous factors. *Cell* **51**, 435–443.
 37. Mueller, P. R. and Wold, B. (1989) In vivo footprinting of muscle specific enhancer by ligation-mediated PCR. *Science* **246**, 780–786.
 38. Pfeifer, G. P., Steigerwald, D., Mueller, B., Wold, B., and Riggs, A. D. (1989). Genomic sequencing and methylation analysis by ligation-mediated PCR. *Science* **246**, 810–813.
 39. Hornstra, I. K. and Yang, T. P. (1992) Multiple in vivo footprints are specific to the active allele of the X-linked human hypoxanthine phosphoribosyltransferase gene 5' region: implications for X chromosome inactivation. *Mol. Cell. Biol.* **12**, 5345–5354.
 40. Manfield, I. and Stockley, P. G. (1994) Ethylation interference, in *Methods in Molecular Biology*, Vol. 30: *DNA-Protein Interactions*. (Kneale, G. G., ed.), Humana, Totowa, NJ, pp. 125–139.
 41. Chalepakidis, G. and Beato, M. (1989) Hydroxy radical interference: a new method for the study of protein-DNA interaction. *Nucleic Acids Res.* **17**, 1783.
 42. Hayes, J. J. and Tullius, Th. D. (1989) The missing nucleoside experiment: a new technique to study recognition of DNA by protein. *Biochemistry* **28**, 9521–9527.
 43. Schickor, P., Metzger, W., Werel, W., Lederer, H., and Heumann, H. (1990) Topography of Intermediates In Transcription Initiation of *E. coli*. *EMBO J.* **9**, 2215–2220.
 44. Shaw, P. E. Schroter, H., and Nordheim, A. (1989) The ability of a ternary complex to form over the serum response element correlates with serum inducibility of the human c-fos promoter. *Cell* **56**, 562–572.
 45. Phillips, S. E. V., Manfield, I., Parsons, I., Davidson, B. E., Rafferty, J. B., Somers, W. S., Margarita, D., Cohen, G. N., Saint-Gorins, I., and Stockley, P. G. (1989) Cooperative tandem binding of Met repressor from *E. coli*. *Nature* **341**, 711–715.
 46. Bushman F. D., Anderson, J. E., Harrison, S. C. and Ptashne, M. (1985) Ethylation interference and X-ray crystallography identify similar interactions between 434 repressor and operator. *Nature* **316**, 651–653.
 47. Newman, P. C., Nwosu, V. U., Williams, D. M., Cosstick, R., Seela, F., and Connolly, B. (1990) Incorporation of a complete set of deoxyadenosine and thymidine analogues suitable for the study of protein-nucleic acid interactions into oligodeoxynucleotides. Application to the EcoRV restriction endonuclease and modification methylase. *Biochemistry* **29**, 9891–9901.
 48. Gogos, J. A., Tzertzinis, G., and Kafatos, F. C. (1991) Binding site selection analysis of protein-DNA interactions via solid phase sequencing of oligonucleotide mixtures. *Nucleic Acids Res.* **19**, 1449–1453.
 49. Gogos, J. A. and Kafatos, F. C. (1994) Determination of sequence preferences of DNA binding proteins using pooled solid-phase sequencing of low degeneracy oligonucleotide mixtures, in *Methods in Molecular Biology*, Vol. 30: *DNA-Protein Interactions*. (Kneale, G. G., ed.), Humana, Totowa, NJ, pp. 295–312.
 50. Blackwell, T. K. and Weintraub, H. (1990) Differences and similarities in DNA binding preferences of Myo-D and E2A protein complexes revealed by binding site selection. *Science* **250**, 1104–1110.
 51. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510.
 52. Taylor, I. A., Watts, D., and Kneale, G. G. (1993) Substrate recognition and Selectivity in the Type IC DNA modification methylase M.Eco R124I. *Nucleic Acids Res.* **21**, 4929–4935.
 53. Taylor, J. D., Ackroyd, A. J., and Halford, S. E. (1994) The gel shift assay for the analysis of DNA-protein interactions, in *Methods in Molecular Biology*, Vol. 30: *DNA-Protein Interactions*. (Kneale, G. G., ed.), Humana, Totowa, NJ, pp. 263–280.
 54. Turner, G. P. and Kneale, G. G. (1993) Cloning, expression and in vitro characterisation of the M13 gene 5 protein. *Biochim. Biophys. Acta* **1173**, 201–208.
 55. Carpenter, M. L. and Kneale, G. G. (1994) Analysis of DNA-protein interactions by intrinsic fluorescence, in *Methods in Molecular Biology*, Vol. 30: *DNA-Protein Interactions*. (Kneale, G. G., ed.), Humana, Totowa, NJ, pp. 313–326.
 56. Schuck, P. and Minton, A. P. (1996) Kinetic analysis of biosensor data: elementary tests for self-consistency. *TIBS* **21**, 458–460.
 57. Taylor, I. A. and Kneale, G. G. (1994) A competition assay for DNA binding using the fluorescence probe ANS, in *Methods in Molecular Biology*, Vol. 30: *DNA-Protein Interactions* (Kneale, G. G., ed.), Humana, Totowa, NJ, pp. 327–338.
 58. Carpenter, M. L. and Kneale, G. G. (1991) Circular dichroism and fluorescence analysis of the interaction of Pfl gene 5 protein with poly(dT). *J. Mol. Biol.* **217**, 681–689.
 59. Carpenter, M. L. and Kneale, G. G. (1994) Circular dichroism for the analysis of DNA-protein interactions, in *Methods in Molecular Biology*, Vol. 30: *DNA-Protein Interactions* (Kneale, G. G., ed.), Humana, Totowa, NJ, pp. 339–346.

60. Taylor, I. A., Davis, K. G., Watts, D., and Kneale, G. G. (1994) DNA binding induces a major structural transition in a type I methyltransferase. *EMBO J.* **13**, 5772–5778.
61. Webb, M., Taylor, I., Firman, K., and Kneale, G. G. (1995) Probing the domain structure of the type IC DNA methyltransferase M.EcoR124I by limited proteolysis. *J. Mol. Biol.* **280**, 181–190.
62. Taylor, I., Webb, M., and Kneale, G. G. (1996) Surface labelling of the type I methyltransferase M.EcoR124I reveals lysine residues in the specificity subunit that are critical for DNA binding. *J. Mol. Biol.* **258**, 62–73.
63. Kneale, G. G. (ed.) (1994) *Methods in Molecular Biology, Vol. 30: DNA-Protein Interactions*. Humana, Totowa, NJ.
64. Busby, S., Kolb, A., and Minchin, S. (1994) Assays for transcription factor activity, in *Methods in Molecular Biology, Vol. 30: DNA-Protein Interactions* (Kneale, G. G., ed.), Humana, Totowa, NJ, pp. 397–411.