

Theory, Characterization, and Modeling of DNA Binding by Regulatory Transcription Factors

Important issues

- *Transcription factors recognize their DNA sites using a variety of different mechanisms.*
- *Parameters besides the DNA recognition sequence regulate binding in vivo.*
- *DNA binding is measured and quantitated using several simple assays.*
- *Modeling DNA–protein interactions is necessary for understanding the mechanism of DNA binding.*

INTRODUCTION	434
CONCEPTS AND STRATEGIES	436
General theory and examples of DNA–protein interactions	436
<i>Theory of DNA recognition, 436</i>	
<i>Chemical basis of the interactions, 437</i>	
<i>The role of the α-helix in DNA recognition, 437</i>	
<i>Major and minor groove specificity, 439</i>	
<i>Monomers and dimers: energetic and regulatory considerations, 441</i>	
<i>Dissociation constant analysis (Box 13.1), 444</i>	
<i>K_d determination, 447</i>	
Analysis and modeling of DNA–protein interactions	448
<i>Identification of a high-affinity DNA recognition site, 448</i>	
<i>Basic theory, 449</i>	
<i>General methods (Boxes 13.2 and 13.3), 449</i>	
<i>Minor groove/DNA backbone probes (Box 13.4), 454</i>	
<i>Major groove probes, 458</i>	
<i>Modeling DNA–protein interactions, 459</i>	
Analysis of promoter-specific multicomponent nucleoprotein complexes	463
<i>DNA binding cooperativity, 465</i>	
<i>DNA looping and bending, 466</i>	
<i>Mechanisms of DNA bending, 468</i>	
<i>Approaches for studying bending, 469</i>	

TECHNIQUES	472
Protocol 13.1 DNase I footprinting	472
Protocol 13.2 Hydroxyl-radical footprinting	482
Protocol 13.3 Phosphate ethylation interference assay	485
Protocol 13.4 Methylation interference assay	488
Protocol 13.5 Electrophoretic mobility shift assays	493
Protocol 13.6 Preparation of ³² P-end-labeled DNA fragments	497

INTRODUCTION

A mechanistic analysis of a promoter generally involves experiments to determine how sequence-specific transcriptional regulatory proteins recognize and bind to their DNA sites, both alone and in combinations. Chapter 9 discussed the criteria for determining the physiological relevance of a DNA–protein interaction. Here we discuss the theory of DNA recognition, how to identify a high-affinity recognition site for a DNA-binding protein, and finally, how to study and model a DNA–protein interaction using chemical and nuclease probes. We then elaborate on some simple principles and methods for studying the formation of multi-activator complexes or enhanceosomes.

The current model is that a eukaryotic DNA-binding protein binds to its physiological sites by continually colliding with nuclear DNA until it encounters a functional site within a promoter. A functional site is one that mediates the physiological action of a transcription factor in the context of a regulated promoter. Nonspecific sites, on the other hand, comprise random sequence or, as we discuss below, specific recognition sequences in an incorrect context. The cell has devised three strategies to expedite and confer specificity to the search. First, much of the untranscribed DNA in a cell is packaged into chromatin, which is largely inaccessible to the regulatory molecule. Second, the concentration of regulatory molecules in the nucleus is raised sufficiently high to overcome any significant competition from nonspecific DNA (i.e., most activators and repressors appear to be expressed at levels of 1000–50,000 molecules/nucleus, likely in excess of their specific sites; this issue is discussed in Ptashne 1992). Finally, a substantial amount of the binding energy is derived from protein–protein interactions that occur only in the proper promoter context.

Context-dependent DNA–protein and protein–protein interactions are central to locating a site. There are two issues that must be considered. First, the actual number of sites which an activator can recognize in naked genomic DNA far exceeds the number of physiological sites. Imagine, for example, a factor that recognizes and contacts a 6-bp site. Statistically, this 6-bp site is present 732,422 times in the human genome (i.e., 3 billion bp divided by 4⁶, the number of combinatorial possibilities for a 6-bp site). Because there are only 50,000–100,000 protein-coding genes in the cell, and it is unlikely that any regulatory factor, with the exception of the general machinery, binds to all of them, the actual number of recognition sites for any given factor very likely exceeds the number of physiologically relevant or functional sites. Second, transcription factors often fall into families that recognize related or identical sites *in vitro* (see Luisi 1995). Because the carefully orchestrated action of transcription factor family members on distinct promoters is critical to the proper functioning and development of eukaryotic cells, these issues raise the question of how physiological specificity is imparted on a DNA–protein interaction.

The enhanceosome theory has been invoked to explain how a protein is able to achieve the proper specificity (see Chapter 1; Echols 1986; Grosschedl 1995; Carey 1998). Figure 13.1 shows the prototypic IFN- β enhancer complex and schematically illustrates its docking with the general machinery. The concept is that the arrangements of sites within a promoter/enhancer and the specific repertoire of regulatory proteins that bind these sites generate a unique network of protein–DNA and protein–protein interactions. The energy or stability of the final structure is dependent on the accurate placement of binding sites and binding of the correct regulatory factors to these sites. The ultimate goal is to assemble a stable complex with the lowest free energy, much like the assembly of a puzzle from its component pieces. In the case shown, the c-Jun/ATF heterodimer binds cooperatively with IRF-3, IRF-7, and NF- κ B to generate an enhanceosome complex.

This view would clarify how an activator distinguishes its physiological sites from non-physiological sites. First, by cooperating with other proteins in a complex as in Figure 13.1A, an activator has a higher affinity for its physiological sites. Presumably, under physiological conditions, only the combination of factors shown in Figure 13.1A could bind and assemble the enhanceosome due to the correct balance of activator concentration and protein–protein interactions. Second, it solves the paradox of related sequence preferences. Although several regulatory proteins may recognize an identical sequence, the subsequent stereospecific protein–protein interactions and the final free energy of the complex would “select” for the correct factor. Another level of selectivity is that the enhanceosome itself would generate a surface complementary to a surface on the Pol II general machinery; only when the correct interface was formed would the enhanceosome loop out the DNA and recruit Pol II, coactivators, and the general factors to the promoter (Fig. 13.1B). Indeed, as discussed in Chapter 1, under such a mechanism the general machinery would assist in assembling the enhanceosome via reciprocal cooperative interactions.

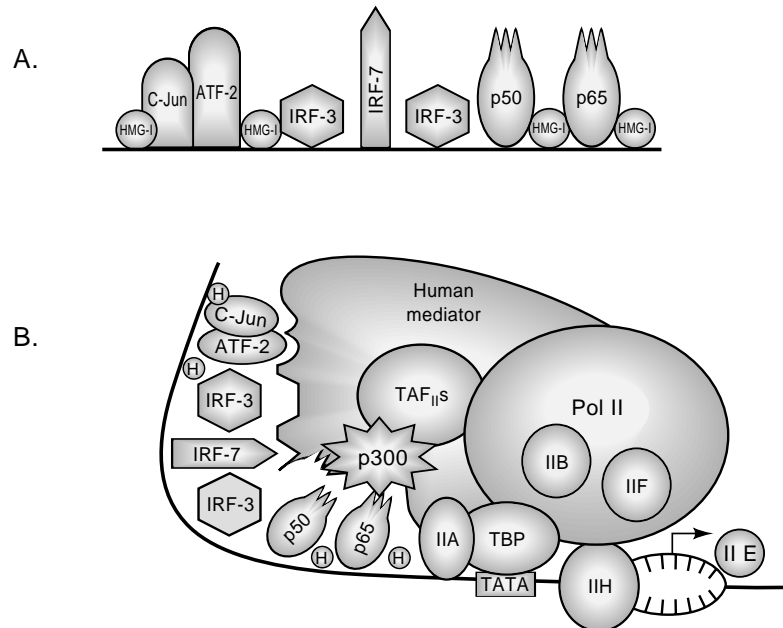


FIGURE 13.1. (A) Schematic of the IFN- β enhanceosome. (Adapted, with permission, from Carey 1998 [Copyright 1998, Cell Press].) (B) Docking with general machinery. (Adapted, with permission, from Ptashne et al. 1998 [Copyright 1998, Elsevier Science].)

According to the hypothesis above, cooperative, promoter context-dependent interactions are the driving force behind distinguishing a functional site from a nonphysiological site. Proteins that deform or bend the DNA, such as HMG-I in Figure 13.1B, may be a necessary component of such complexes. These proteins could permit certain combinatorial protein-protein interactions otherwise restricted by the limited flexibility of the intervening DNA within its persistence length and the size and flexibility of the bound proteins.

This chapter examines biochemical methods and strategies for understanding basic aspects of promoter recognition. We initially focus on how a single cloned regulatory protein recognizes a site, followed by a summary of methods for studying DNA bending and cooperative binding, two phenomena necessary to generate more sophisticated and specific enhanceosome complexes.

CONCEPTS AND STRATEGIES

General Theory and Examples of DNA-Protein Interactions

Theory of DNA Recognition

DNA site recognition by a regulatory protein is generally influenced by both specific interactions with the bases and nonspecific interactions with the phosphate/sugar backbone. Although there is some mild controversy surrounding the use of this “simplified” terminology to describe DNA binding, it nevertheless provides a framework that can be refined on the basis of the regulatory context.

In a typical interaction the regulatory protein (P) and DNA site (S) are in reversible equilibrium with the protein-site (PS) complex. The equilibrium is represented as $PS \leftrightarrow P + S$ with an equilibrium dissociation constant of $K_d = k_1/k_2 = [P][S]/[PS]$. This equilibrium constant, also defined as the ratio of the forward (k_1) and reverse (k_2) rate constants, takes into account all of the enthalpic and entropic energies contributing to binding, including the cost of locating the site ($\Delta G = \Delta H - T\Delta S$). This K_d can also be defined in terms of free energy, using the Gibbs free-energy equation $\Delta G = -RT \ln K_d$ (for a discussion of how this equation bears on biological reactions, see Dill 1997).

The energy and specificity of a protein-DNA interaction are generated by a unique stereospecific array of amino acid side chains that are chemically and spatially complementary to an array of chemical groups displayed by the bases in the major or minor groove of the DNA. Each chemical interaction provides a quantum of free energy; each deviation from the ideal site generates a dramatic reduction in site affinity due to the logarithmic relationship between free energy (ΔG) and K_d as described above by the Gibbs equation. This logarithmic relationship is one mechanism for enhancing specificity. The form of complementarity described above is called direct readout.

Another form of DNA recognition is called indirect readout, and it concerns the ability of a protein to bind a specific sequence based on the DNA secondary structure, or conformation. If the recognition site deviates from the optimum, the inherent deformability of the DNA at the site may be affected. The resulting change in K_d may be much greater than would be predicted by the loss of energy from simple chemical interactions. The *EcoRI* GAATTC recognition site, for example, depends on a specific sequence array to accommodate a deformation. Substitution of a single base within this site alters the ΔG of binding and subsequently raises the K_d (Lesser et al. 1990).

In addition to the specific interactions of amino acid side chains with the exposed groups of base pairs, nonspecific interactions with the relatively uniform phosphate backbone of the B-DNA helix also contribute to the binding energy or K_d . Crystallography studies suggest that these interactions often provide the bulk of the free energy within the

recognition complex, although they do not generate specificity (Pabo and Sauer 1992).

In addition to the noncovalent enthalpic effects described above, entropic contributions derived from the release of ordered water and salt ions from a site upon binding are believed to be significant driving forces in protein–DNA interactions (Ha et al. 1989).

Studies on the partitioning of energy in site-specific DNA recognition have not yet yielded a satisfactory general understanding of the problem. However, significant advances and a renewed interest in the chemistry and physics of site recognition suggest a solution in the future.

Chemical Basis of the Interactions

We now examine more closely the primary enthalpic contributions to DNA binding—the interactions between the amino acid side chains, and occasionally, backbone amide and carbonyl groups, with both the phosphate/sugar backbone and the exposed chemical groups of base pairs displayed in the major or minor grooves. The four classes of specific interactions in the major and minor groove are:

1. The C5-methyl group of thymines and the C5-hydrogen group of cytosine participate in van der Waals contacts with the aliphatic amino acids (Fig. 13.2A).
2. Specific H-bonds with the exposed edges of base pairs and the phosphates along the DNA helix (Fig. 13.2B, C). Amino acid–DNA H-bonding interactions are supported by a variety of amino acid side-chains (Fig. 13B) and both the amide and carbonyl groups of the peptide backbone (Fig. 13C). All specific protein–DNA complexes employ this strategy.
3. Water molecules can serve as a hydrogen-bonding bridge between an amino acid and a base pair or phosphate (e.g., the *EcoRI* and Trp-R co-complexes: Otwinowski et al. 1988; Narayana et al. 1991; Shakked et al. 1994) (Fig. 13.2D). This concept has arisen as a result of higher-resolution protein–DNA structures.
4. Occasionally an amino acid side chain will intercalate between two bases. Observed mainly in DNA-bending proteins, this interaction is exemplified by the TBP, which inserts two phenylalanines between the first and last base pairs of the TATA box. The intercalation generates an 80° bend, unwinds the helix, and widens the minor groove. HMG proteins employ a similar strategy (Werner and Burley 1997). We cover this topic in more detail later in the chapter.

The Role of the α -Helix in DNA Recognition

How does a protein specifically recognize DNA? The size and shape of the α -helix (cylindrical; main chain is 4.6 Å, in diameter) are ideal for fitting into the major groove (helical diameter, 19 Å; major groove rise, 17 Å) and, not surprisingly, the vast majority of DNA-binding proteins have employed this strategy (Fig. 13.2E). The chemical diversity and flexibility of the amino acid side chains and the rotation of the helical axis endow the α -helix with a large number of potential recognition surfaces for binding a specific DNA sequence. Crystal studies also show that the disposition of this so-called recognition α -helix in the groove relative to the backbone axis of the DNA varies extensively among different regulatory proteins. However, because there is little evidence that an isolated α -helix is capable of independent recognition (Pabo and Sauer 1992), it has been proposed that interactions with the phosphate backbone, mediated by other protein elements, are essential for properly positioning the helix on its site. Thus, once on its site, the recognition helix is stabilized by a protein scaffold (including adjacent α -helices) and an intricate network of

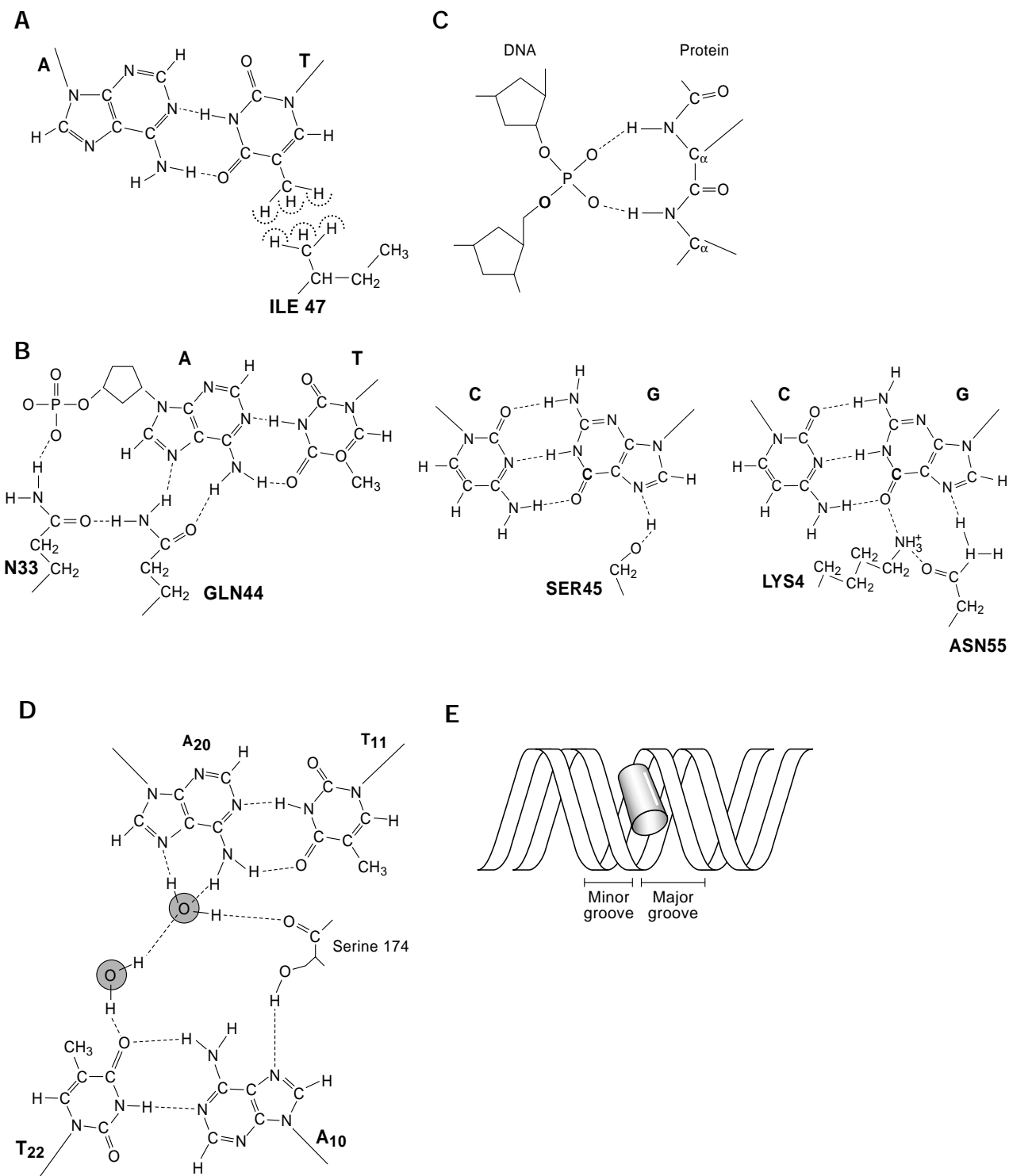


FIGURE 13.2. Chemical basis of interactions for DNA binding. (A, Adapted, with permission from Kissinger et al. 1990 [Copyright Cell Press] 1992.) (B, Adapted, with permission, from Branden and Tooze 1991 [Copyright 1991 Garland Publishing].) (C, Redrawn, with permission, from Jordan and Pabo 1988 [Copyright 1988 American Association for the Advancement of Science].) (D, Adapted, with permission, from Feng et al. 1994. [Copyright 1994 American Association for the Advancement of Science].) (E) An α -helix in major groove. (Adapted, with permission from Ptashne 1992 [Copyright Blackwell Science].)

nonspecific interactions with the DNA backbone. Despite the predilection of structures employing α -helices in the major groove, there are many examples of β -sheets used in DNA recognition, including the prokaryotic Met and Arc repressors (Somers and Phillips 1992; Raumann et al. 1994; Suzuki 1995) and the eukaryotic p53 protein (Cho et al. 1994). Furthermore, many proteins use, in addition to major groove interactions, flexible stretches of peptide that reach back and make contacts into the minor groove, thereby enhancing specificity.

Major and Minor Groove Specificity

We have described the energetics of DNA site recognition above, but we have not discussed what constitutes the complementary surface on DNA that permits it to dock with a protein surface in a sequence-specific fashion. That is, how do different DNA sequences mark themselves to be identified by a regulatory protein? The conformation or shape of the DNA, although displaying minor sequence-dependent changes in its dimensions and torsion angles, is nonetheless relatively uniform in structure, as are the phosphate backbone and sugar conformations. Instead, specificity is imparted by the sequence-dependent projection of chemical groups from the bases into the major and minor grooves.

When A and T, or C and G, hydrogen-bond to form the base pairs connecting the antiparallel strands of the DNA, all four bases expose chemical groups that are either multivalent or not engaged in pairing. These exposed groups could conceivably bond with amino acid side chains or the peptide backbone on the recognition surface of the regulatory protein. These chemical groups, however, could only impart specificity if the arrangement of accessible H-bond donors and acceptors in the major and minor groove changed dramatically with the DNA sequence. In 1976, Seeman and colleagues (Seeman et al. 1976) considered this issue by examining the arrangements of chemical groups in different base pairs. They concluded, for reasons that are described below, that each base pair displays a unique three-dimensional pattern of chemical groups in the major groove. A multibase recognition site displays a combinatorial and, hence, more elaborate array. Thus, the uniqueness of an array increases with the size of the site. This variation is most evident in the major groove and less evident in the minor groove.

How does the array vary? There are four combinations of base pairs—A:T, T:A, G:C, and C:G. Each base pair, in addition to containing H-bonded chemical groups that permit formation of the base pair, also displays certain unbonded chemical groups that fall into three categories: (1) the H-bond acceptor (ac); (2) the H-bond donor (do), or (3) the van der Waals contacts with a C5 hydrogen on cytosine (vdw-h) or the methyl group projected from the C5 of thymine (vdw-me) (Fig. 13.3). In the case of 5'AT3', the major groove displays the spatial array ac/do/ac/vdw-me, whereas TA displays vdw-me/ac/do/ac. Similarly, a GC base pair displays ac/ac/do/vdw-H, whereas a CG base pair displays vdw-h/do/ac/ac. Thus, each base pair is unique with respect to the array of chemical groups. In contrast, the minor groove of a TA base pair displays ac/ac, whereas an AT displays ac/ac as well. A GC displays ac/do/ac, whereas a CG displays ac/do/ac as well. Thus, there is significant diversity in the major groove, and AT can be distinguished from TA, and GC from CG; the minor groove, on the other hand, is relatively bare and AT cannot be distinguished from TA or GC from CG, only AT/TA from GC/CG. For this reason, as well as the accessibility issue, the major groove appears to be the primary target for sequence specificity. This may be oversimplified because studies using substituted pyrroles revealed subtle differences in recognition of A-T versus T-A base pairs in the minor groove. These differences have not yet been observed in crystal structures of DNA-protein complexes (Kielkopf et al.

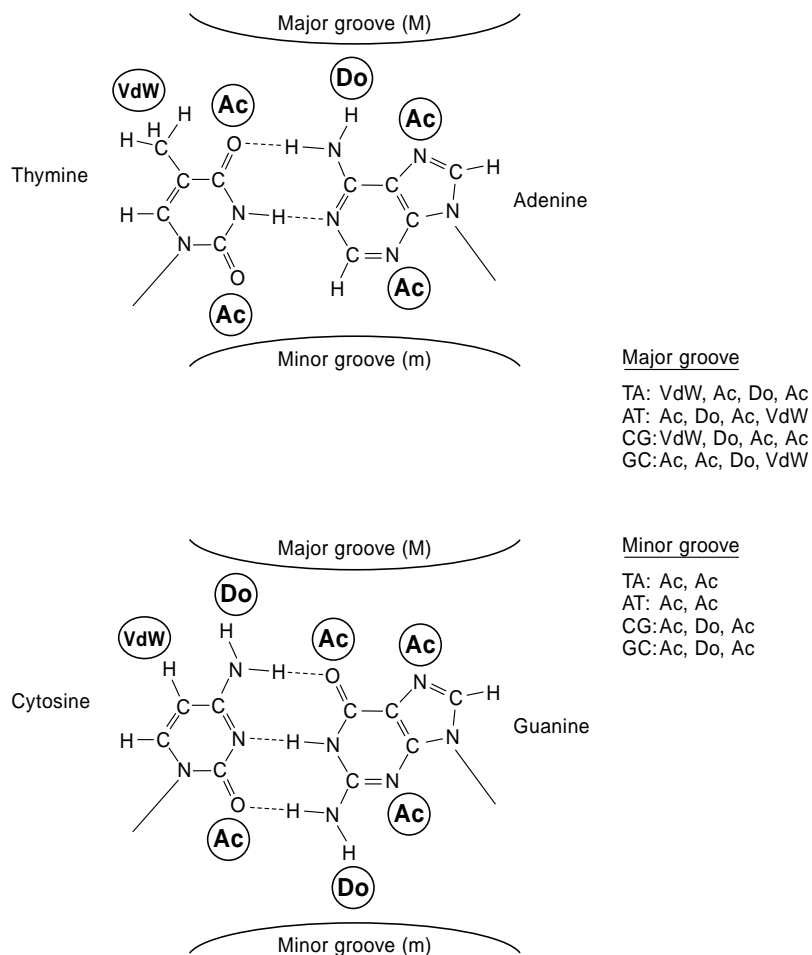


FIGURE 13.3. H-bond donors and acceptors (Adapted, with permission, from Watson et al. 1987 [Copyright 1987 Benjamin Cummings Publishing].)

1998) but are likely to occur as the crystal and nuclear magnetic resonance (NMR) solutions of additional minor groove binding proteins are published.

Despite the predilection that most DNA-binding proteins have for the major groove, well-described cases of specific minor groove recognition do indeed exist. As described above, some proteins couple major and minor groove recognition, whereas others interact predominantly with the minor groove. Examples of the coupling approach include the Oct-1 POU and MAT α -2 homeodomains, among others. These contact the major groove with the homeodomain helix-turn-helix or HTH motif and the minor groove with an extended amino-terminal arm (see, e.g., Wolberger et al. 1991; Klemm et al. 1994). In prokaryotes, the Hin recombinase also employs minor-groove as well as major-groove contacts (Feng et al. 1994). Some proteins recognize the minor groove exclusively. A more unusual mechanism of such minor groove specificity involves the binding of TBP to the TATA box (for a review, see Burley and Roeder 1996). TBP employs 10 anti-parallel β -sheets to form a concave undersurface that recognizes and binds in the minor groove of the TATA site, a point we cover later in this chapter. HMG domain proteins also employ extensive contacts with the minor groove, as discussed below and later in the chapter. LEF-1, an HMG-box protein, distorts and bends the DNA helix at the minor groove to facilitate the formation of an enhanceosome at the TCR- α enhancer. The HMG domain is an L-

shaped molecule that molds the minor groove to accommodate the three α -helices (Grosschedl et al. 1994; Love et al. 1995). Both TBP and the HMG domain employ intercalating amino acids as part of their mechanism of specificity. Taken together, these examples show that minor groove binding is an important component for sequence recognition by several major classes of DNA-binding proteins.

Currently there is no universal code for the recognition of DNA sequences by proteins. The amino acid side chains or main-chain amido and carbonyl groups engage in sequence-specific interactions with a wide variety of bases or with the phosphate backbone. However, there may be a preference for forming hydrogen bonds with purines, and there is some degree of conservation in sequence recognition within families of DNA-binding proteins, as within the family of homeodomains or zinc fingers (for review, see Pabo and Sauer 1992). A considerable amount of research has been applied to devising zinc fingers with altered specificities (Rebar and Pabo 1994), and the results of these studies may soon reveal some general rules for amino acid–base pair interactions for at least one class of transcription factors (Pomerantz et al. 1995). Further research into DNA-binding interactions may yet reveal more discernible patterns for DNA sequence recognition.

Monomers and Dimers: Energetic and Regulatory Considerations

Repetition of a DNA recognition unit is one of the most widely employed strategies used by nature to design DNA-binding proteins (Fig. 13.4A). There are several ways in which this repetition is employed: (1) dimerization or formation of another higher-order oligomer (Fig. 13.4AII) and (2) multimerization of a DNA recognition unit (Fig. 13.4AIII). Because the affinity is exponentially related to the free energy of binding ($\Delta G = -RT \ln K_d$), doubling the binding energy by doubling the number of recognition units leads to an exponential increase in affinity of a dimer versus a monomer, or a monomer bearing tandem recognition units versus a monomer with a single unit. Dimers like the yeast GAL4 and GCN4 proteins bind to 17- and 7-bp sites, respectively, each site displaying pseudo-twofold rotational symmetry (Carey et al. 1989; Oliphant et al. 1989; Ellenberger et al. 1992). Crystal structures have revealed that the proteins bind DNA with each monomer recognizing one-half of the rotationally symmetric sites.

Many proteins increase their regulatory diversity by heterodimer formation, with each monomer recognizing one of the half-sites. This strategy has two purposes. It allows heterodimers to recognize a site bearing nonsymmetric half-sites, because each monomer has a different sequence preference. Alternatively, both partners in the heterodimer may have the same DNA-binding specificity but have unique regulatory properties. Excellent examples of the former include the binding of RXR/retinoic acid receptor (RAR) heterodimers to an asymmetric direct repeat (DR) on a promoter; RXR recognizes the 5' half-site, and RAR (or another partner) recognizes the 3' half-site. Depending on RXR's partner, the heterodimer complex can regulate transcription from several different DRs (for review, see Mangelsdorf and Evans 1995) and respond to different combinations of ligands. The Max protein product can also bind to a number of related proteins. Myc/Max and Mad/Max heterodimers bind a conserved 6-bp regulatory site, but the heterodimers have very different regulatory effects. The Myc/Max dimer activates cell-cycle-dependent genes, whereas Mad/Max dimers repress these same genes. Another example is the eukaryotic Jun protein, which can bind DNA either as a homodimer (Jun-Jun) or as a heterodimer (Jun-Fos, Jun-CREB). Jun homodimers bind weakly to the AP-1 promoter element, and Jun-Fos heterodimers bind tightly. Jun heterodimers with other family members also bind to a wide array of sites (for an older but insightful review, see Herschman 1991).

Multimerization of the DNA-binding motif within a single polypeptide is another powerful approach used to reinforce specificity. Multimeric proteins can contain either multiple identical recognition units or several units with distinct modular structures. The zinc-finger proteins like Zif 268 are a good example of the first case. In Zif 268, three C-C-H-H fingers recognize similar tandem sequence motifs on the double helix. The zinc fingers are well-designed to position the recognition helix of Zif 268 into the major groove by forming a “C-” shaped structure around the double helix (Pavletich and Pabo 1991). POU-domain homeobox proteins like Oct-1 employ the second method of binding-domain multimerization. Oct-1 contains two specific DNA recognition units, the POU homeodomain and the POU-specific domain. The POU homeodomain is a 3-helix subunit where helices 2 and 3 form the helix–turn–helix (HTH) (see Chapter 1) motif and minor-groove binding occurs via the amino-terminal arm. The POU-specific domain contains four α -helices, which employ extensive base and phosphate-backbone contacts with the double helix. These two domains together bind an 8-bp site, and both domains are required for efficient binding (Klemm et al. 1994).

Strategies for determining the oligomeric state. Determination of the oligomeric state of a protein, either in solution or when it binds to its site, is a relatively simple task. There are four general methods:

1. *Chemical and nuclease footprinting techniques.* These techniques reveal whether a dimeric protein binds to a symmetrical site. The approach can also indicate whether symmetrically opposed mutations in the dyad alter binding. We elaborate on these methods in the second section of this chapter.
2. *Heterodimer analysis.* This strategy requires that one have in hand a set of deletion derivatives that bind to the site and give rise to complexes with unique electrophoretic mobilities in electrophoretic mobility shift assay (EMSA) analyses. Two such derivatives are chosen (often they can be synthesized by *in vitro* transcription and translation) and assayed individually or in combination for the ability to bind the site. The association of monomers in an oligomer follows a binomial distribution $(a+b)^n$ where a is the concentration of one derivative, b is the concentration of the other, and n is the oligomeric state. If the protein were a dimer, then the distribution of EMSA complexes would be predicted by the equation, $a^2+2ab+b^2$. Thus, each derivative alone would generate one complex as shown in Figure 13.4B, and together these derivatives would form a new complex that migrated with intermediate mobility. If a and b were present at equimolar concentrations, the complexes would be present in a 1:2:1 distribution. Whatever the concentrations of a and b , they would be factored into the equation to predict the distribution. If the protein were a trimer, the distribution would be predicted by $(a+b)^3$.

There are two problems with this analysis. If the proteins contain a strong dimerization interface, the two monomers may not exchange upon mixing to form heterodimers, thereby confounding the analysis. One way to circumvent this problem is to use a variety of salt and buffer conditions, because some conditions may promote intersubunit exchange. Another more common option is to use *in vitro* transcription and translation to cosynthesize the two derivatives. Thus, the two synthetic RNAs can be combined and cotranslated in the same mixture, allowing the subunits the opportunity to associate as they are synthesized.

Figure 13.4C shows an example of an actual experiment using GAL4 derivatives. Two different GAL4 derivatives were synthesized individually, or cotranslated, incubated with a labeled 17-mer GAL4 site and fractionated on nondenaturing polyacrylamide gels. An autoradiograph of the gel is shown. Each derivative gave rise to a unique shifted com-

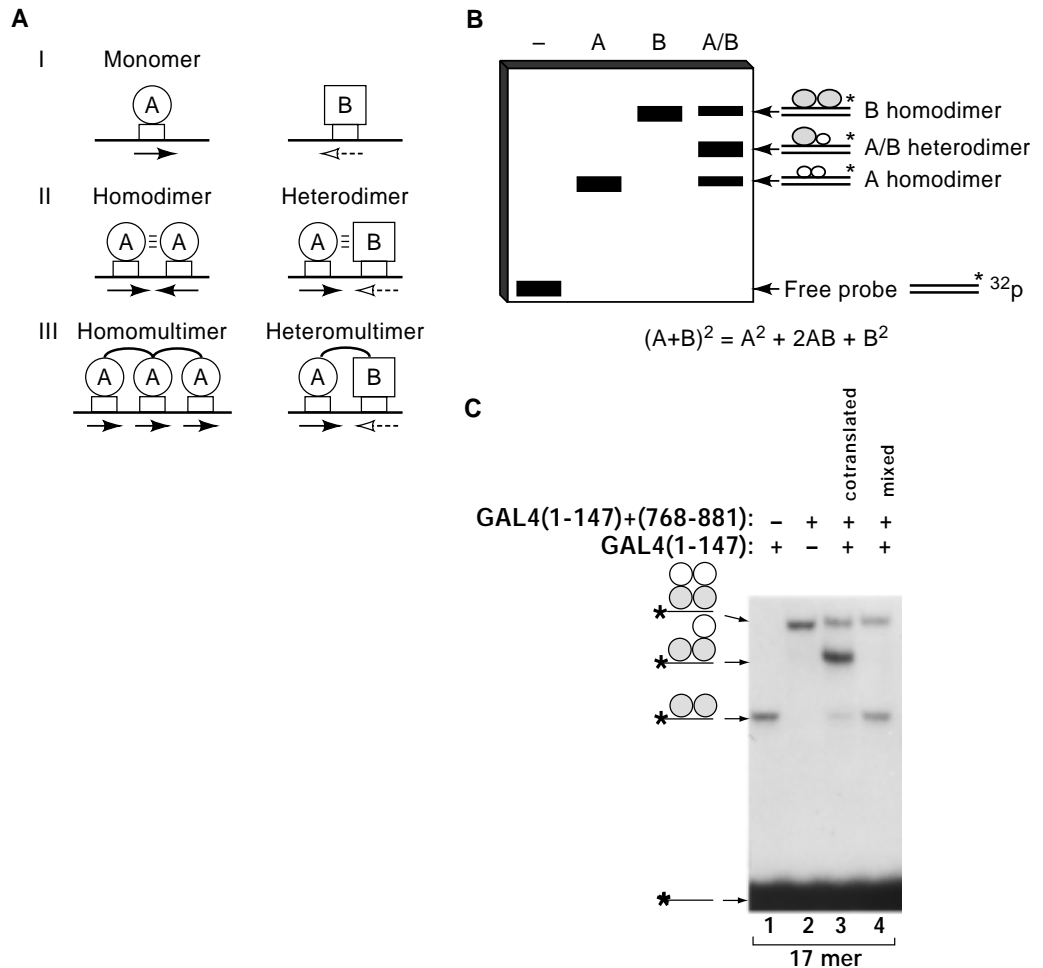


FIGURE 13.4. Heterodimer analysis. (A) Different strategies for repetition of a DNA-binding unit. (B) Theory. (C) GAL4-derivatives experiment. (Adapted, with permission, from Carey et al. 1989 [Copyright Academic Press Ltd.])

plex (lanes 1 and 2), whereas the two derivatives together generated a new shifted complex with intermediate mobility (lane 3). The appearance of a single new complex was evidence for heterodimer formation and demonstrated that GAL4 bound its site as a dimer, as illustrated by the schematics next to the autoradiogram. However, mixing the two derivatives together did not generate the intermediate shifted complex, demonstrating that the GAL4 proteins had strong dimer interfaces and the subunits could not exchange once formed (lane 4).

3. *Chemical crosslinking followed by SDS-gel electrophoresis.* Glutaraldehyde is a common agent used in crosslinking experiments (it crosslinks primary amines). Different concentrations of recombinant regulatory factors are incubated with increasing amounts of glutaraldehyde and then, after quenching with primary amines (Tris, ethanolamine, etc.), the products are fractionated on SDS-polyacrylamide gels against untreated protein. The shift of the protein from its predicted molecular weight to a higher-molecular-weight species is generally indicative of higher-order oligomer formation; but be aware that glutaraldehyde is relatively nonspecific, thus controls like BSA must be used

in conjunction. Measuring dimer formation at concentrations of protein used for EMSA and in the presence and absence of DNA can also be helpful. Such information may help determine the mechanism of dimerization.

4. *Gel filtration resins.* Gel filtration chromatography is a relatively outdated method that can separate the monomeric and dimeric species on the basis of molecular mass. The approach is less reliable because the shape of the protein can dramatically influence the mobility of a protein in gel filtration columns. An accurate measure of molecular weight depends on a combination of gel filtration and sedimentation velocity experiments. The apparent molecular weight of the complex in such experiments can be compared with the molecular weight of the monomer to determine the oligomeric state.

Dissociation Constant Analysis

Because the biological activity (activation or repression) of a recognition site is often proportional to its affinity for a regulatory protein, affinity is a useful criterion for comparing proteins binding to related but distinct binding sites. The affinity of a protein for its site is generally expressed in terms of a dissociation constant or K_d . The K_d is an equilibrium constant that describes the ratio of reactants to products. K_d is expressed as a molar concentration. Typically, the K_d varies depending on the temperature, buffer, salt conditions, and method of measurement. The variability is unique to each protein and, hence, absolute measures of K_d for two different proteins are generally not comparable unless they were performed under the same conditions and preferably side-by-side. K_d values for DNA-binding proteins are measured by nitrocellulose filter binding, EMSA, or DNase I footprinting, although recent technological advances have resulted in machines such as the Pharmacia Biacore that use surface plasmon resonance (SPR) to measure K_d . SPR and other physical methods for K_d measurement, such as fluorescence anisotropy, are becoming more common in the gene expression field.

One common misconception is that the ratio of protein to DNA is a meaningful parameter in determining the K_d . The ratio is actually meaningless except under very special circumstances, which we will comment on. Instead, the absolute concentrations of protein and DNA, and the conditions under which binding is studied, are the essential parameters for the measurement. Because K_d measurements are becoming more common, we discuss some simple principles to consider when performing such an analysis. We begin with an analysis of a simple equilibrium between a DNA site (S) and an activator protein (P). The reaction is as follows: Activator [P] binds to its DNA site [S] and forms an activator:DNA complex [PS]. This is represented schematically in Equation 1, which by convention is most often illustrated as the dissociation reaction



The algebraic representation of the reaction is shown by Equation 2

$$K_d = \frac{[P][S]}{[PS]} \quad (2)$$

There are three standard scenarios that a researcher encounters when studying DNA-protein interactions:

1. *Subsaturating.* This condition occurs when there may be billions of molecules of S and P in a reaction mixture but the concentrations of both are well below the K_d . Under such conditions, even if equivalent numbers of molecules are present, the components are not at high enough concentrations to form an appreciable amount of complex (<1%). In an EMSA reaction, no shifted complex would be observed.

2. *One-component saturating.* When one component, P for example, is equal to or greater than the K_d while S is limiting, then P will begin to saturate S. In an EMSA, 50% of the probe would be saturated when P is equal to the K_d and, hence, 50% of the DNA would be shifted to the bound form. In such a case, while S was nearly saturated there could still be many molecules of unbound P.
3. *Stoichiometric interactions.* When both components are present at concentrations exceeding the K_d , they interact largely in a stoichiometric fashion. Thus, if P and S were present at equal concentrations, they would always exist in the form of a complex, in contrast to the subsaturating case.

Examples of the latter two scenarios are explained in Box 13.1.

Box 13.1.

Examples of K_d Scenarios

We examine the most common scenario, scenario 2, first and then scenario 3. Let us assume for a moment that the K_d of P for its site is 10^{-9} M. Let us now measure the occupancy of S when we set the P protein concentration equal to the K_d (10^{-9} M) and the S at the following two concentrations lower than the K_d : (A) 2×10^{-11} M and (B) 2×10^{-15} M. Because the K_d is a constant, the DNA and protein concentrations must be manipulated using approximations to come to a final tally equal to the K_d . The amounts of protein and DNA have been purposely arranged to minimize the manipulations. You would obtain the following result in case A (Eq. 2a): (Remember Eq. 2a is the same as Eq. 2 with the values filled in)

$$K_d = 10^{-9} = \frac{[10^{-9}][10^{-11}]}{[10^{-11}]} = \frac{[P][S]}{[PS]} \quad (2a)$$

Note that the amount of P that binds S (10^{-11}) is so negligible that it does not effectively change the concentration of free protein (10^{-9} M), and so an *approximation* of the concentration of free protein can be used in place of algebra (i.e., 10^{-9}). An exact concentration of the free protein can be calculated using algebra, but it would have very little effect on the overall result. Also note that we started with 2×10^{-11} M DNA, which is partitioned between S and PS.

$$\frac{[S]}{[PS]} = \frac{10^{-11}}{10^{-11}} = 1 = 50\% \text{ occupancy of DNA} \quad (3)$$

Because the amounts of bound DNA (PS) and free DNA (S) are equivalent (Eq. 3), then we say that the DNA is occupied 50% of the time. Hence, the operational definition of K_d —the concentration of free protein for which 50% of the DNA is bound when the DNA concentration is limiting (i.e., well below the K_d). The percentage of bound or complexed DNA is often referred to as the fractional saturation or occupancy.

We now examine case B. It is evident that if we adjust the DNA to 2×10^{-15} M, as shown below (Eq. 4), we observe the same fractional occupancy as when the DNA was 2×10^{-11} M. Because the concentration of protein that binds to the DNA is negligible, the ratio of the bound and unbound DNA remains 1, which is equivalent to 50% occupancy (Eq. 5).

$$10^{-9} = \frac{[10^{-9}][10^{-15}]}{[10^{-15}]} = \frac{[P][S]}{[PS]} \quad (4)$$

$$\frac{[S]}{[PS]} = \frac{10^{-15}}{10^{-15}} = 1 = 50\% \text{ occupancy of DNA} \quad (5)$$

The relevant points are: (1) The protein concentration can be thousands of fold greater concentration than the DNA, yet the DNA is only partially occupied. (2) When the DNA concentration is below the K_d , at any given protein concentration the same fractional occupancy of the sites will be observed.

Using this set of conditions, where the DNA concentration is below the K_d , one can plot either the S/PS (right axis) or the percent saturation of a probe (left axis) as a function of protein concentration expressed on a log scale. Figure 13.5 shows the parabolic increase in saturation as the protein concentration is raised. Note that in the simple bimolecular interactions described above, to change probe saturation from 10% to 90% or from S/PS of 9 to S/PS of 0.11 entails an 81-fold increase in protein concentration.

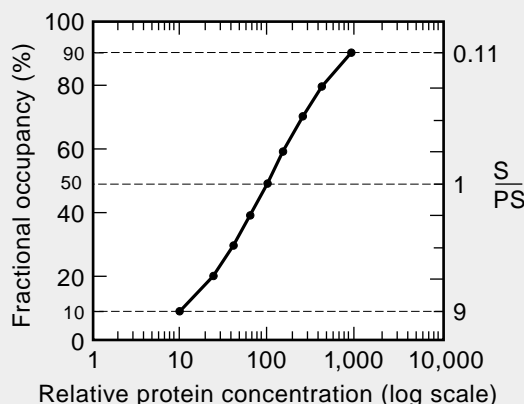


FIGURE 13.5. Binding isotherm for simple bimolecular reaction.

A different type of result from that described above is observed when the DNA concentration is near to or greater than the K_d (scenario 3). If the protein concentration added to the reaction is left at 10^{-9} and the DNA concentration is adjusted to 10^{-8} (Eq. 6), the following results are obtained using approximations. (You could also raise the protein instead. The K_d does not usually refer to any one component of the reaction.)

$$10^{-9} = \frac{[10^{-10}][10^{-8}]}{[10^{-9}]} = \frac{[P][S]}{[PS]} \quad (6)$$

The result is that almost all of the starting protein enters into a complex with the DNA. The overall free DNA concentration, however, does not change because it is much greater than the protein concentration.

Take another example. The starting protein and DNA concentrations added to the reaction are now raised well above the K_d to 10^{-7} for each. $[P]=10^{-7}$ and $[S] = 10^{-7}$

$$10^{-9} = \frac{[10^{-8}][10^{-8}]}{[10^{-7}]} = \frac{[P][S]}{[PS]} \quad (7)$$

To balance the amounts of P, S, and PS, almost all of the protein and DNA must interact in a stoichiometric fashion. Thus, almost all of the 10^{-7} M added DNA and protein form the PS complex. A smaller amount (10^{-8} M using approximations) remains in the form of free P and S. Although this latter scenario is commonly viewed to be occurring during standard DNA–protein interaction studies, in fact, usually one of the components is limiting and the situations described in Equations 2a and 4 apply. The scenario, however, is a common way to determine the activity of a DNA-binding protein as described in the text.

K_d Determination

There are several well-established methods for determining K_d , many of which employ EMSA or quantitative DNase I footprinting analysis. One of the most traditional approaches relies on employing the concept depicted in Equation 7: When the protein and DNA concentrations are raised above the K_d , the interaction becomes stoichiometric. Therefore, if the DNA concentration is known, the protein concentration can be determined empirically. The method relies on three conditions: (1) a protein that is sufficiently pure to accurately determine its absolute concentration by protein determination assays such as the Bradford, Lowry, or biuret methods; (2) the availability of ample amounts of an oligonucleotide bearing a high-affinity site; (3) an assay that can accurately measure binding of the protein in solution.

If EMSA is to be employed, preliminary experiments should be performed to demonstrate that the EMSA and DNase I footprinting assays generate similar results. The EMSA is more quantitative because it is quite easy to measure binding by densitometric scanning of the unbound and bound DNAs. However, occasionally the EMSA conditions are not always optimal and can favor dissociation of the protein–DNA complexes during the running of the gel. An effect called caging, where the limited volume of the gel pores minimizes the dissociation or promotes rapid reassociation, may also influence the results.

The first step is to determine the apparent K_d of a protein by titrating the protein with a very low amount of ^{32}P -labeled oligonucleotide ($\sim 10^{-11}$ M). The concentration of protein that is required to generate 50% occupancy is known as the apparent K_d . The term “apparent K_d ” is used because although the amount of protein is known, the amount of *active* protein is not. Many recombinant proteins, particularly those requiring special structures (e.g., zinc nucleated protein folds), are not entirely active due to partial denaturation or oxidation of key residues (e.g., cysteines) during purification.

To ensure that the DNA is indeed limiting (i.e., below the K_d) in the initial titration, it should be varied fivefold in either direction, and the fractional occupancy (50%) should remain the same. This concept is illustrated in Equations 2a and 5. Once the apparent K_d is known, the protein is raised 50- to 100-fold above the apparent K_d . This will lead to almost complete occupancy of the probe. At this stage the precise amount of active protein is unknown, but it is known that levels of the active protein have been raised well above the K_d into the range where it will interact stoichiometrically with the DNA. The next step is to gradually add unlabeled competitor DNA of known concentration. As the competitor oligonucleotide is raised, it begins to compete with the bound ^{32}P -labeled oligonucleotide for the protein. Because the DNA concentration exceeds the K_d , when the competitor oligonucleotide begins to compete (i.e., reduced occupancy in the EMSA), it is assumed that it must be interacting stoichiometrically with the protein. Eventually, all of the active protein in solution is quantitatively sequestered by the oligonucleotide, and the amount of oligonucleotide bound is identical to the amount of active protein.

Therefore, when the competitor begins to exceed the amount of active protein, unbound radiolabeled DNA is observed either by EMSA or footprinting. When 50% of the labeled oligonucleotide is competed to the unbound form, the oligonucleotide competitor would have exceeded the amount of active protein by twofold. The active protein concentration is then calculated to be equivalent to one-half of the oligonucleotide concentration. Because the molar concentration of oligonucleotide is known, the active protein concentration can be easily calculated. Once the active concentration is known, one can return to the original experiment and replace the apparent K_d with a true K_d . The method is facile and can be performed without any special algebraic manipulations. The advantage of this

approach is that if the protein of interest is an oligomer, a dimer for example, the high concentrations of protein used in the stoichiometry method will favor the dimer species. Thus, one determines the active amount of dimer in the reaction mixture.

Note that because the free DNA is always related to the bound DNA by the K_d equation, oligonucleotide competition can be used to calculate the K_d at any concentration of protein. However, the algebraic manipulations are more complicated and, depending on the approach, make several assumptions regarding protein activity or the oligomerization status. One disadvantage of performing a measurement when the protein concentration is near or below the K_d is that, when the protein is an oligomer, unless the oligomerization constant is lower than the K_d , a large portion of the protein is in the monomeric form. Since only the dimer will be effectively competed by unlabeled oligonucleotide, this scenario could in principle result in a lower value for the K_d than that determined using the stoichiometry method.

Analysis and Modeling of DNA–Protein Interactions

Fortunately, due to the rapid advances made in solving the crystal or nuclear magnetic resonance (NMR) structures of DNA-binding proteins and their complexes with DNA, our knowledge of DNA recognition is quite advanced. Generally, knowing that a protein falls into a recognizable family for which a crystal structure is available can instantaneously provide a general overview of the sequence preference of a regulatory factor and provide an outline for understanding how the protein recognizes DNA.

Nuclease and chemical probing can provide invaluable information about how a protein recognizes its site. It can be used to refine a model deduced for a related family member, test the validity of a crystal or NMR structure, and explore the binding of previously uncharacterized regulatory protein families. In the following sections we discuss the information that can be obtained by such approaches. We first cover the approach necessary to identify a site if one has a DNA-binding protein but a high-affinity recognition site is not known.

Identification of a High-affinity DNA Recognition Site

In many cases, the investigator has a protein in hand because it was isolated from genetic screens to identify important regulatory proteins or perhaps scored as an oncogene. There may be some evidence that the protein binds DNA, but its sites of action are not known. Proteins such as Myc, MyoD, and orphan nuclear receptors are classic examples of this scenario. In such cases, how does one identify a site to initiate mechanistic studies on DNA binding?

In *Saccharomyces cerevisiae* and *Drosophila melanogaster*, organisms with facile genetics, many regulatory proteins have been cloned through genetic screens, and their promoter sites have been identified and confirmed by promoter mutagenesis of affected genes. Inactivating either the site or the factor genetically permits an assessment of whether the factor is acting through the site *in vivo*. Studies in mammalian systems, where there is not a facile genetic system, generally rely on a myriad of approaches that couple mutagenesis with either transfection or biochemical assays. The strategies listed below describe approaches for the identification of a high-affinity binding site necessary to initiate binding studies. The approaches do not describe how to determine whether that site is a physiological recognition sequence associated with a naturally responsive promoter. This issue is covered in detail in Chapter 9.

1. *Promoter scanning.* This is scanning the promoter of a responsive gene by DNase I footprinting and EMSA analysis using overlapping restriction fragments as probes.
2. *Selection techniques.* These approaches, which fall into several categories, are the preferred methods for identification of a high-affinity site. In one, the DNA-binding protein is coupled to a solid affinity matrix and a mixture of random DNAs (either oligonucleotides, random restriction fragments, or fragments encompassing a responsive promoter) is passed through. Fragments that are retained can be reselected and then subcloned for biochemical analysis (Kalionis and O'Farrell 1993). A variation of this technique is selected and amplified binding site (SAAB) analysis (Blackwell 1995). The SAAB method employs a binding technique such as a solid affinity matrix or EMSA to identify DNA fragments or oligonucleotides that bind a protein. The DNA fragments are then purified and amplified by PCR and reselected. By varying the stringency of the binding reaction, high-affinity sites can be obtained for binding analyses. A variation of the method has also been called SELEX (systematic evolution of ligands by exponential enrichment).
3. *One-hybrid analysis.* An alternative genetic technique that has proven successful has been to use yeast as a tool to identify the binding site (described in Chapter 8). In these cases, the random DNA fragments or suspected binding sites from promoters or enhancers are cloned upstream of a yeast reporter gene comprising a core promoter linked to β -gal on a yeast vector, preferably integrated into the genome. The transcription factor is then introduced into yeast under control of a strong promoter (i.e., ADH). In some cases, a yeast activation domain is fused to the mammalian DNA-binding domain. This approach is the basis of the yeast one-hybrid assay for identifying promoter sites and DNA-binding domains from cDNA libraries (Wang and Reed 1993).

Basic Theory

All nucleolytic and chemical probing methods are based on the “nested set” theory used to describe DNA sequencing (see Maxam and Gilbert 1977). A nested set contains a unique end and a variable end (Fig. 13.7). When DNA molecules containing a unique end marked by an end label, usually ^{32}P , and a variable end represented by a modification or cleavage site, are fractionated by denaturing gel electrophoresis, the size of the labeled single-stranded DNA fragment, as measured on a gel, will be a precise measure of the distance to the variable end, i.e., its position relative to the unique end. The principle of sequencing and enzymatic and chemical probing is to produce DNA fragments containing a unique end and a variable end positioned at a specific base. The length of the radioactive single-stranded DNA molecules on a gel then specifies the precise position(s) of that base. In most footprinting methods, the unique end is generated by directly ^{32}P -end-labeling the DNA fragment.

General methods

In initiating any modeling analysis of a DNA–protein interaction, one begins with a simple approach. After the recombinant protein is purified, a DNA fragment bearing the site of interest is generated. It is best to use small DNA fragments whose banding pattern on a sequencing gel can be resolved to the base pair. Although synthetic oligonucleotides are generally adequate for measuring binding in gel retardation assays, they are, for reasons that will be elaborated on below, only suitable for studying binding of a small protein to a short site. Therefore, the binding site is generally first subcloned into an easily manipula-

Box 13.2

SAAB Analysis

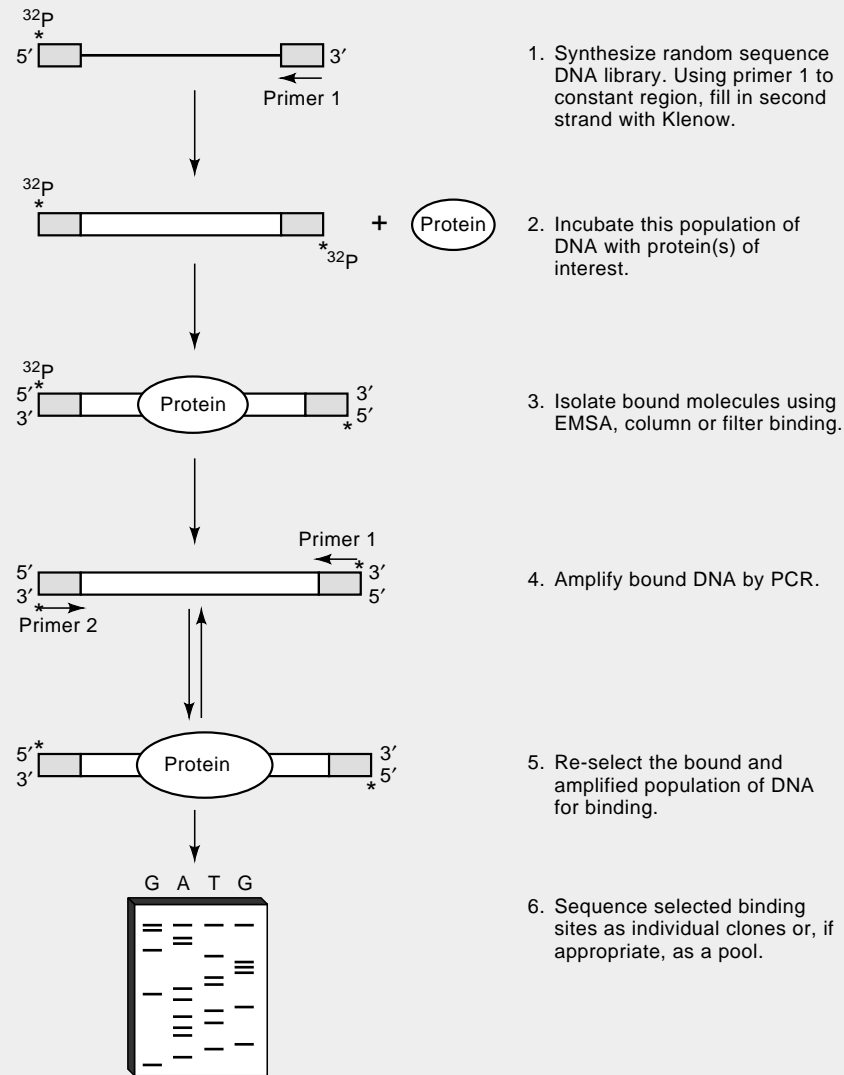


FIGURE 13.6. Flowchart for SAAB analysis (Adapted, with permission, from Blackwell 1995 [Copyright Academic Press].)

SAAB methodology (Fig. 13.6) was developed by Blackwell and Weintraub (Blackwell 1995) for comparing binding sites for homo- and heterodimers of the bHLH proteins MyoD and E2A. An independent but conceptually similar approach was developed by Ellington and Szostak (1990) to select RNA molecules with unique binding properties. The method, particularly when applied to RNA, has been termed SELEX in subsequent publications (Tuerk et al. 1992). The SAAB approach has since been used to identify the binding-site repertoire for dozens of regulatory proteins. The approach is imaginative and simple, employing PCR technology coupled with EMSA. An oligonucleotide that contains a random sequence flanked by constant PCR sites is constructed and ^{32}P end labeled. A second oligonucleotide complementary to the constant region is synthesized and annealed to the first oligonucleotide. The annealed oligonucleotide is then used as a primer for Klenow DNA polymerase, which is used to fill in the gap

to generate a double-stranded molecule. The ^{32}P -labeled molecule is then incubated with saturating or subsaturating concentrations of the protein; the complexes are separated from unbound DNA by gel mobility analysis. The bound DNA containing the selected sequence is reamplified by PCR and subjected to one or more reselection steps. One advantage of this approach is the ability to manipulate the binding conditions and the number of selection cycles to distinguish high-affinity from low-affinity sites.

Note that the selection technique need not employ EMSA. Chromatography of the DNA over solid affinity matrices such as glutathione-*S*-transferase (GST)-transcription factor fusions attached to glutathione Sepharose or immunoprecipitation of the DNA-protein complexes with an antibody to the transcription factor are alternative methods of identifying the bound sites.

ble vector. Optimally, for sites in the range of 7 to 17 bp, the site is centered within a 50-bp region that can be either excised as a uniquely ^{32}P -end-labeled restriction fragment or prepared by PCR using a primer pair where only one primer is ^{32}P -labeled. After fragment purification and subsequent enzymatic or chemical analysis of the protein-DNA complex, the products are fractionated on a 7–10% polyacrylamide gel. This manipulation generates a cleavage ladder of sufficient detail for careful and quantitative analysis.

The two main approaches that give a broad indication of the binding of a protein to its site are DNase I (Galas and Schmitz 1978) and exonuclease III footprinting (Box 13.3). Due to the large size of the nucleases and steric blockage by the DNA-bound protein, the footprint is often considerably larger than the recognition site itself, and the precise edge of the binding site is difficult to locate. Nevertheless, these methods provide a necessary starting point for studying and modeling any DNA-protein interaction. There are several methods that employ small chemicals such as copper-phenanthroline or methidium-propyl EDTA that bind in the minor groove and cleave the DNA (Box 13.3). These reagents generally allow a more precise definition of the site borders (Sigman et al. 1991).

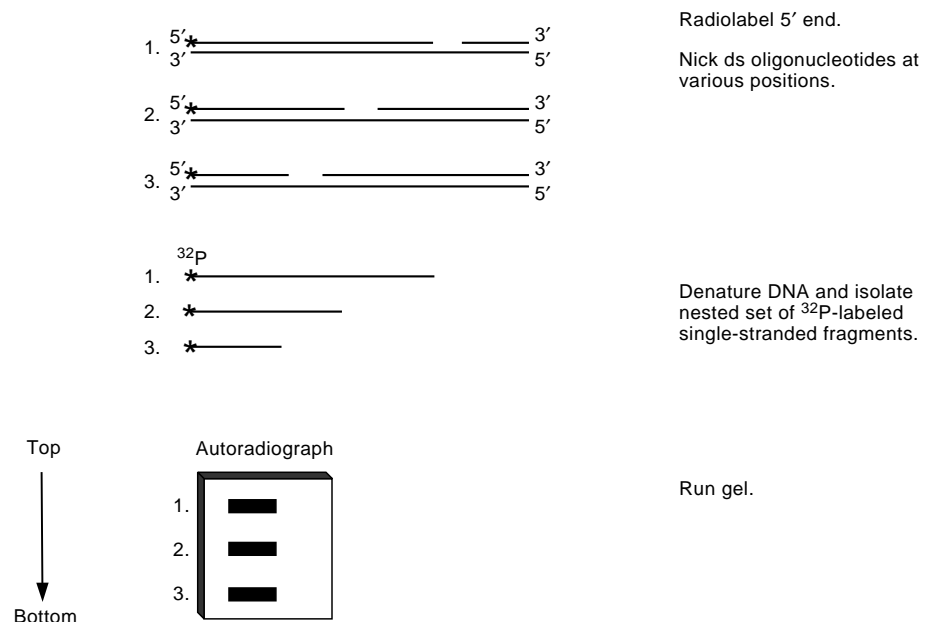


FIGURE 13.7. A nested set of oligonucleotides.

Box 13.3

DNase I and Exonuclease Footprinting

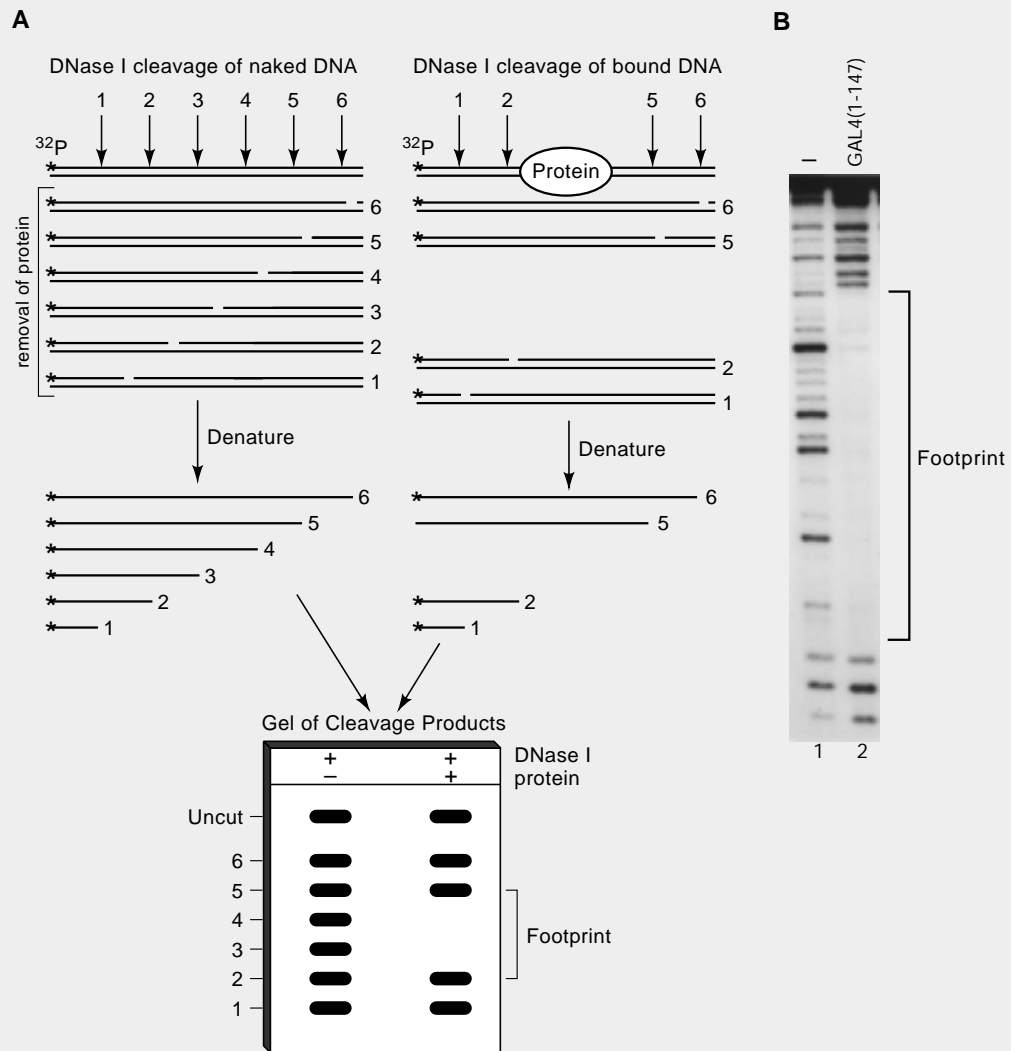


FIGURE 13.8. (A) DNase I footprinting theory. (B) Example of GAL4 DNA-binding domain (Adapted, with permission, from Carey et al. 1989 [Copyright Academic Press Ltd.])

DNase I footprinting (Fig. 13.8) is one of the most widely used methods to study binding by a regulatory protein to its recognition sequence (see Protocol 13.1) (Galas and Schmitz 1978). DNase I interacts with the DNA in the minor groove. Although it binds and cleaves the minor groove in a relatively nonspecific fashion, its activity is dependent on certain sequence-based structural characteristics that the binding site adapts (Suck 1994). Local structural deformities within the site, like differences in propeller twist or local minor groove width, can significantly influence the binding and/or cleavage properties of DNase I.

In the presence of Ca^{++} and Mg^{++} , DNase I cleaves a single strand of the double-stranded DNA (manganese ions can lead to double-stranded cleavages) and generates a 5'- PO_4 and 3'-

OH products. Bound protein protects the DNA site from cleavage. For many proteins, the DNase I footprint is significantly larger than the recognition sequence (possibly because the DNase I protein itself is so bulky), and hence it is considered a low-resolution probe. The footprints are often staggered because DNase I is asymmetric in shape, and on a particular strand, will more closely approach the edge of the site on one side than the other. Figure 13.8A illustrates the footprinting theory, and Figure 13.8B shows an example using the GAL4 DNA-binding domain. It is relatively important in footprinting to cleave the DNA on average only once per molecule to observe a digestion ladder. This will entail only partial cleavage of the probe. DNase I will not cleave where protein is bound and the missing cleavage products form a “footprint on the gel.”

Exonuclease III (*ExoIII*) is a 28-kD nuclease isolated from *E. coli*. It preferentially binds to a free 3' hydroxyl on double-stranded DNA and then cleaves inward in a semiprocessive fashion. It prefers a recessed or blunt 3' hydroxyl group for binding, releases nucleoside monophosphates, and possesses a 3' phosphatase activity. *ExoIII* is employed by molecular biologists, in conjunction with single-stranded endonucleases, to create deletions of genes or their promoter regions (see Chapter 7). However, it can also be used to localize protein-binding sites (Siebenlist et al. 1980). The enzyme cleaves until it encounters the binding site, whereupon it halts (Fig. 13.9). Generally, given enough time, the nuclease can wait until the protein falls off its site and can pass through it, resulting in “read-through” on the autoradiograph. *ExoIII* footprints are slightly smaller than DNase I footprints. For GAL4, *ExoIII* generates a 21-bp footprint, whereas with DNase the footprint is 27 bp (Carey et al. 1989).

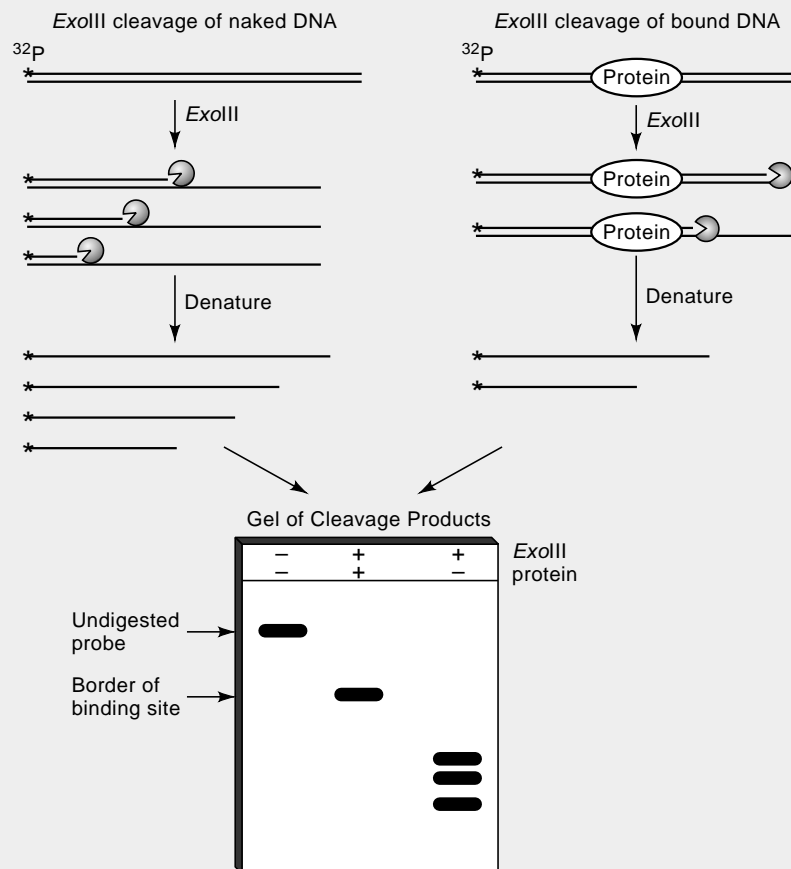


FIGURE 13.9. Exonuclease III footprinting.

Other general footprinting reagents

MPE and Cu-phenanthroline footprinting. Although DNase I and ExoIII footprinting provide a broad indication of site binding, smaller chemical cleavage reagents can define the edges of the site more accurately, as well as provide important structural information on the interaction. Methidium propyl EDTA (MPE) and Cu-phenanthroline (OP-Cu) are two chemical cleavage reagents that can provide this type of information. Both reagents bind in the minor groove; MPE intercalates between the base pairs. MPE generates a diffusible hydroxyl radical that cleaves the DNA at the sugar/phosphate backbone. OP-Cu is believed to cause a more directed cleavage by attacking the C-1 hydrogen on the sugar. OP-Cu can be an especially useful reagent because the nicking assay can be carried out within an acrylamide gel on otherwise unstable protein–nucleic acid complexes (Sigman et al. 1991).

Minor Groove/DNA Backbone Probes

The positioning of a protein along the DNA backbone can provide important information about its mechanism of recognition. As described above, interaction of the protein with backbone phosphates serves an important role both in providing energy of binding and in allowing these interactions to position elements of the protein in the major groove. There are two main techniques for probing minor groove and backbone interactions: hydroxyl radical protection (Protocol 13.2 and Box 13.4) and ethylation interference (Protocol 13.3 and Box 13.4). Unlike DNase I and OP-Cu footprinting, which also bind the minor groove but give a larger picture of both major and minor groove interactions, hydroxyl-radical protection (because the radical is small and diffusible, unlike phenanthroline which must intercalate) and ethylation interference can provide a more detailed view of exclusively minor groove interactions. It is for this reason that these techniques have become so popular in footprinting assays. Hydroxyl radical footprinting identifies interactions near sugar residues, whereas ethylation interference identifies interactions with phosphates directly.

Box 13.4

Chemical Probes for Minor Groove Interactions

The hydroxyl radical attacks many residues of the deoxyribose sugar with a preference for the hydrogens on the deoxyribose ($C5' = C4' > C3' = C2' = C1'$) and abstracts the hydrogen atoms attached to those residues (see Dixon et al. 1991). It was previously thought to be a minor groove reagent, but it is more strictly stated as a probe for contacts near the sugar/phosphate backbone (Balasubramanian et al. 1998). The abstraction of the H atom initiates a series of electron-transfer reactions resulting ultimately in DNA strand scission. A hydroxyl radical molecule is as small as a water molecule and thus is not subject to the same steric restrictions as DNase I. Therefore, hydroxyl-radical footprinting reveals detailed information about primary contacts along the backbone and minor groove. Hydroxyl radical is somewhat insensitive to small perturbations in backbone structure except in the case of kinks and bends where the cleavage efficiencies can vary. Unfortunately, hydroxyl radical also attacks proteins. Some proteins are more sensitive than others to the radical or to the chemical used to generate it. Hydroxyl radicals are typically generated using the Fenton reaction, as shown below (Fig. 13.10A). Figure 13.10B shows a typical hydroxyl-radical footprinting reaction where increasing concentrations of GAL4 DNA-binding domain (lanes 3–5) were incubated with DNA and

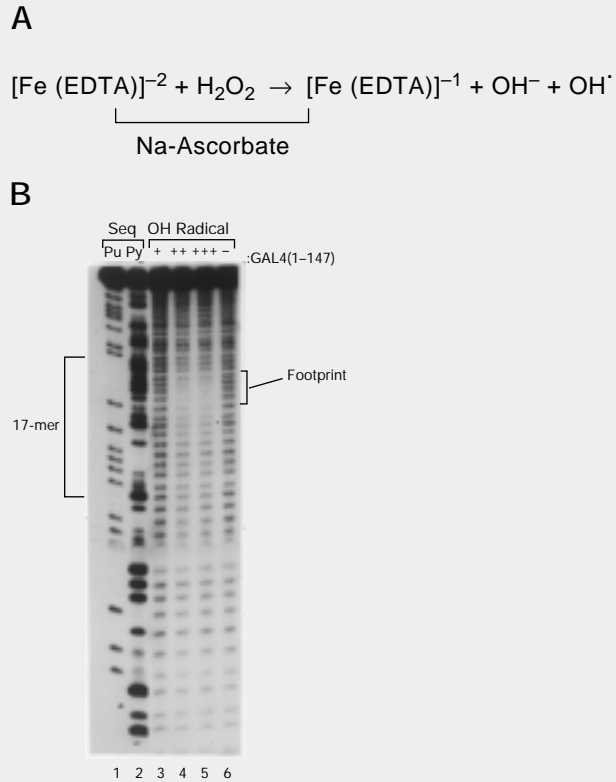


FIGURE 13.10. (A) Fenton reaction. (B) Hydroxyl-radical footprinting reaction of GAL4. (Adapted, with permission, from Carey et al. 1989 [Copyright Academic Press Ltd.] .)

cleaved with radical. A small footprint located near the center of the 17mer was observed when compared with DNA alone (lane 6). The position of the footprint was determined by comparing the positions against a DNA sequencing ladder. Lanes 1 and 2 show Maxam-Gilbert purine and pyrimidine cleavage ladders. We will later employ this information in modeling GAL4-DNA interactions (see section on Modeling DNA-protein interactions).

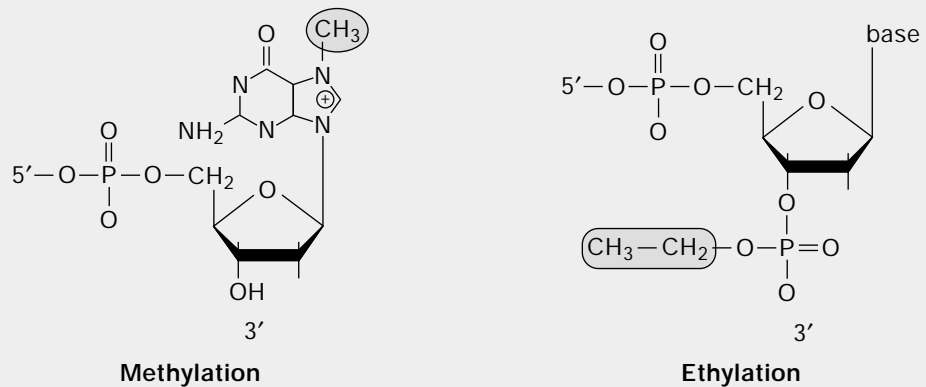


FIGURE 13.11. Methylated Gs and ethylated phosphates.

Ethylation interference relies on the ability of ethylated phosphates to interfere with protein binding (Protocol 13.3) (Siebenlist et al. 1980; Manfield and Stockley 1994). The information content is similar but not identical to hydroxyl-radical footprinting, a point that we return to below. First, the DNA is ethylated with ethylnitrosourea, which modifies primarily phosphates along the backbone (Fig. 13.11). The modification is performed to ensure one ethylation per DNA molecule (Fig. 13.12). By taking into account the Poisson distribution of the process of modification, only about 10% of the molecules should be allowed to become ethylated before a significant amount of molecules are modified twice or more (Fig. 13.12). This can be determined by modifying DNA until approximately 10% of the starting probe is converted to a ladder of bands. It is critical to modify only once so that in the final analysis it is clear that the modification being detected— the one nearest to the ^{32}P -labeled end— is the one responsible for interference. The presence of too many modifications on the same molecule will confound the analysis. This same principle holds for other interference and protection assays.

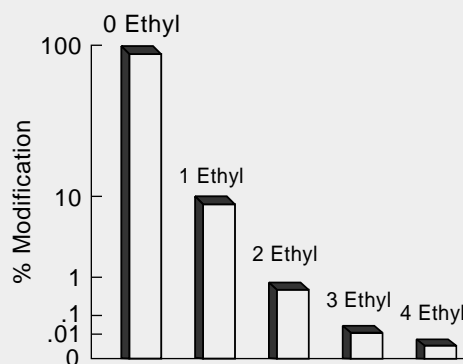


FIGURE 13.12. Optimal Poisson distribution of DNA molecules in different states of ethylation.

After modification, barely saturating amounts of the protein are bound to the DNA (Fig. 13.13A). Barely saturated amounts are used because although the modification decreases binding, it does not always abolish it. High concentrations of protein may overcome the deleterious effect of the modification. The unbound and bound DNA fractions are then separated by EMSA. The bound fraction of DNA is enriched in DNA molecules modified at positions that do not interfere with protein binding and, hence, are not in close proximity to the protein (steps 2 and 3). The unbound fraction is modified at positions that do interfere with protein binding. The points of modification can be identified by isolating the bound and unbound fractions (e.g., bands from an EMSA gel; step 3), treating the DNA with piperidine, which cleaves at the affected residues, and fractionating the cleaved fragments by denaturing gel electrophoresis (steps 4 and 5) alongside a sequencing ladder of the same fragment (Fig. 13.13A). The sequencing ladder can be generated by using the Maxam-Gilbert chemical method or by simply preparing a sequencing primer with the same 5' end as the ^{32}P -label and then using the dideoxy sequencing method.

The data are usually clear in that most positions on a given stretch of DNA will not interfere and, hence, the bound fraction contains a ladder of bands spanning the length of the gel. A few bands will, however, be depleted in the bound fraction and enriched in the unbound fraction. These bands represent modifications that interfere with binding and, thus, positions that come into close proximity with bound protein. The data are generally summarized by superimposing

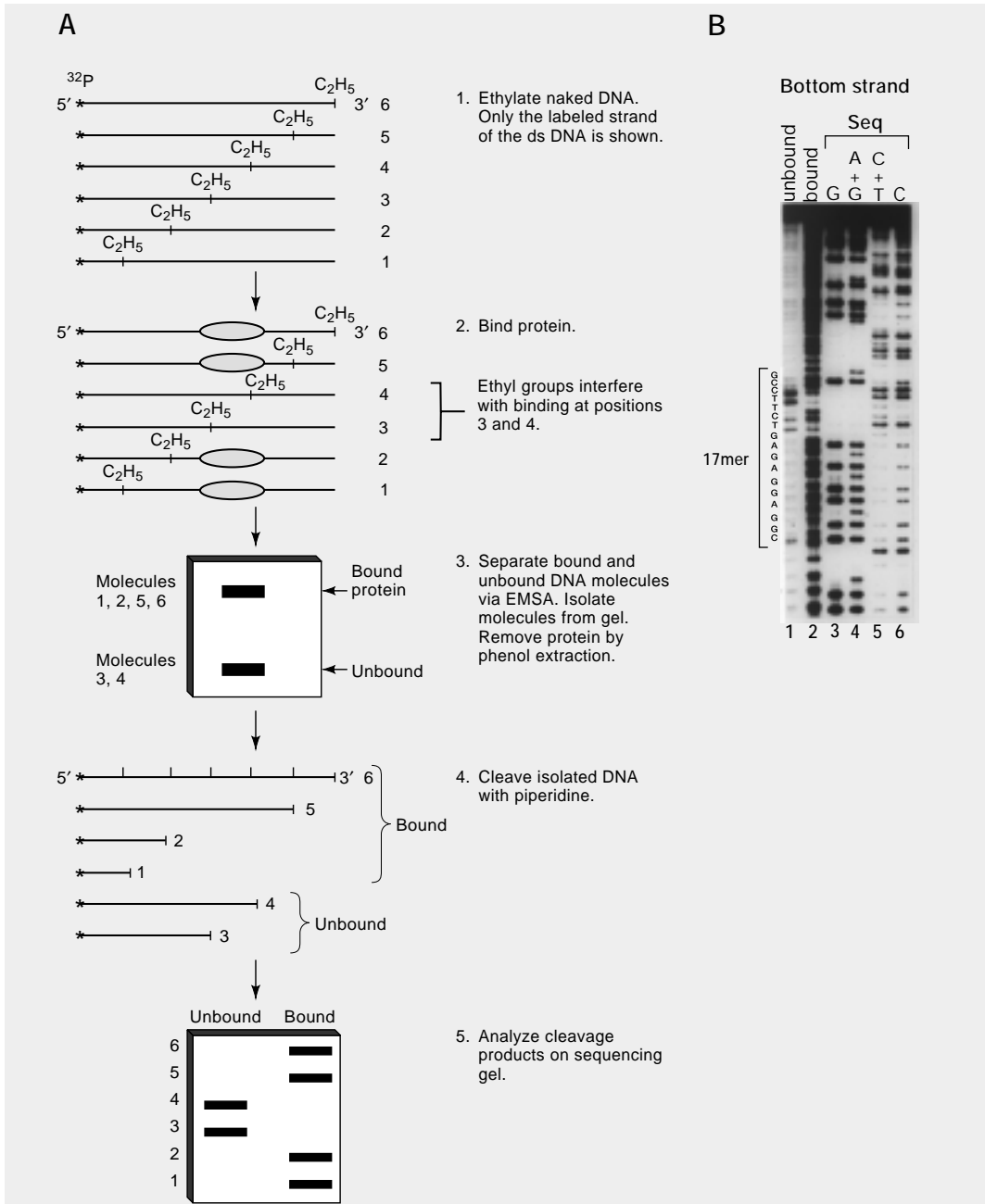


FIGURE 13.13. (A) Ethylation interference. (B) Experiment with GAL4 (Adapted, with permission, from Carey et al. 1989 [Copyright Academic Press Ltd.])

the contacts onto a schematic B-DNA helix. If the protein falls into a family whose structure is known, the data can be modeled using that structure as a starting point.

Figure 13.13B illustrates a typical experiment using GAL4. The bands in the unbound lane represent phosphates that interfere with GAL4 binding, and the bands in the bound lane represent phosphates that do not interfere. Note the depletion of phosphates from the bound, and the enrichment in the unbound, lanes. Positions were determined by a Maxam-Gilbert sequencing ladder. We will later show how this information can be used for modeling GAL4's interaction with its site (see section on Modeling DNA-protein interactions).

Major Groove Probes

There are several approaches to understanding how a protein interacts in the major groove of the DNA. These include dimethyl sulfate (DMS) protection and interference (Siebenlist et al. 1980; Wissmann and Hillen 1991), the missing nucleoside approach (Hayes and Tullius 1989; Dixon et al. 1991), and standard site-directed mutagenesis (see Chapter 8).

DMS methods. The principle of DMS protection is similar to DNase I footprinting, but DMS is a very small reagent and hence can provide detailed information of an interaction. DMS methylates the ring nitrogens N7 of G (major groove) (Fig. 13.11) and N3 of A (minor groove). Unlike DNase I, which generates a protection over the entire binding site, only bases that come into very close proximity to a protein are protected from modification by DMS. Furthermore, bases protected from methylation can often correlate with bases directly involved in an interaction with protein, as deduced from crystal or NMR structures. However, this correlation does not always hold up because protein binding can inhibit methylation of uncontacted bases in a site. Therefore, the chemical probing analysis is often correlated with a mutational analysis of the protein and site. We discuss examples below to demonstrate the limits of these correlations. Some proteins are sensitive to conditions required for DMS protection or to DMS itself, which often modifies proteins, causing them to bind poorly (e.g., TFIIIA; Fairall et al. 1987). In such cases, an alternative is the DMS interference assay.

In DMS interference, the DNA is methylated in the absence of protein on average of once per molecule followed by isolation of the modified DNA. The modified DNA is then interacted with a protein, and the bound and unbound samples are isolated and analyzed. The methodology and DNA processing chemistry are much like ethylation interference. Many investigators proceed directly to DMS interference simply to avoid the potential complications of DMS protection.

Missing nucleoside and mutagenesis. Among the other approaches that can be employed to study major groove interactions, missing nucleoside and mutagenesis (Chapter 8) are the simplest. In mutagenesis, one or more bases in the site are replaced with other bases, and the change, usually a decrease, in affinity is measured. Certain bases play a major role in affinity. Altering these bases removes chemical groups critical to the interaction, which leads to a reduction in affinity.

The missing nucleoside approach is somewhat analogous to an interference experiment (Dixon et al. 1991). The hydroxyl radical is used to remove a nucleoside (base-sugar) prior to interacting with a protein. Afterward, the gapped DNA is incubated with saturating amounts of protein and the bound and unbound fractions are separated by EMSA. Again the bands are excised and electrophoresed on a gel alongside a sequencing ladder. The bound fraction is enriched in molecules missing nonessential bases, whereas the converse is true for the unbound fraction. The information content of this approach is probably greater than with the DMS and ethyl interference techniques, but again, the gapping can alter the structure of the DNA site, which may influence affinity.

Chemical substitution. Base analog or chemical substitution is an excellent, albeit technically difficult, method for determining the effect of substituting certain chemical groups on affinity. In this example, certain functional groups on the base are selectively replaced, keeping the remaining structure of the base intact. K_d analysis is used to measure the consequences. Base analog substitution can minimize the potential structural perturbations (see Lesser 1990). The substitution of inosine for guanosine is an example of using base analog substitution. Such an approach was used to study TBP binding. The thymines and adenines in the TATA box were exchanged with cytosines and inosines, respectively (Starr and Hawley

1991). These changes converted the major groove of TATAAAA to that of the sequence CGCGGGG but did not alter the display of chemical groups in the minor groove (which remained the same as with TATAAAA). The authors showed that the substitution had little effect on TBP binding, demonstrating that TBP was largely binding in the minor groove.

A primer on the energetics of major groove contacts. To understand the energetic consequences of altering an amino acid–base pair interaction via mutagenesis, recall that a typical hydrogen bond imparts 3–7 kcal of energy whereas a van der Waals bond imparts 1–2 kcal. Let's calculate the resulting decrease in affinity due to loss of one hydrogen bond (i.e., via a mutation). Using the Gibbs equation, we can calculate the decrease in affinity for the loss of 1 kcal of energy: if $\Delta G = -RT \ln K$, then at 25°C, $\Delta G = -(1.98)(273+25)2.3 \log K$, with each unit in the standard state. Each 10-fold decrease in affinity is therefore accompanied by a decrease of 1.36 kcal of free energy. Thus, for every loss of 1 kcal in binding energy, the dissociation constant decreases 7.35-fold. Therefore, the loss of one hydrogen bond (~3 kcal) between a base and an amino acid side chain results theoretically in a 22-fold decrease in affinity. One caveat of the mutagenesis approach is that changing the base can also have subtle effects on the overall structure of the site, i.e., different bend, twist, or roll angles which contribute to DNA recognition or deformability of the site (e.g., *EcoRI*). These caveats should be kept in mind when interpreting data, particularly with proteins that bend or kink the DNA.

Modeling DNA–Protein Interactions

To demonstrate how the chemical and nuclease probing technology can be employed to study a protein–DNA interaction, we have chosen two well-characterized examples in which the probing chemistry or “footprint phenotypes” (Yang and Carey 1995) can be compared directly with the protein–DNA structure. The examples discussed will be GAL4 and TBP. GAL4 binds in the major groove while TBP engages in predominantly minor groove contacts. In both cases, essential aspects of the DNA-binding mechanism were deduced from the results of chemical and nuclease protection and later confirmed by the crystal structure of the co-complexes (Carey et al. 1989; Lee et al. 1991; Starr and Hawley 1991; Kim et al. 1993a, b; Marmorstein et al. 1992).

GAL4. We attempt to show here that the crystal structure of GAL4 DNA-binding domain with its 17-bp site bears out many of the predictions made from modeling GAL4–DNA interactions using chemical and nuclease protection. *Figure 13.14 summarizes the domain organization of GAL4 and essential elements of the crystal structure in a graphic form. Fig. 13.15 then compares the contacts made in the structure with those deduced from chemical probes.*

The DNA-binding domain is located within the first 94 amino acids and can be subdivided into a DNA recognition domain between amino acids 1 and 65 and a dimerization domain from 66 to 94 (Fig. 13.14A). Embedded within the DNA recognition domain are 6 cysteines between amino acids 11 and 38 (boxed) that chelate two zinc ions to nucleate folding of the region into a structure that recognizes the GAL4-binding site. The cysteines and their corresponding protein fold are often called a binuclear cluster, and GAL4 was the founding member of a large fungal protein family with this conserved DNA recognition motif (Schwabe and Rhodes 1997; Todd and Andrianopoulos 1997). Although the DNA recognition domain is a monomer in solution, it forms a dimer on DNA. This dimer binds very weakly, and the natural dimerization region, located between amino acids 65 and 94, greatly increases affinity of GAL4 for its site (e.g., see Carey et al. 1989). Nevertheless, when attempting to crystallize GAL4, the fragment bearing 1–65 formed crystals with DNA and its structure was solved in 1992 (Marmorstein et al. 1992).

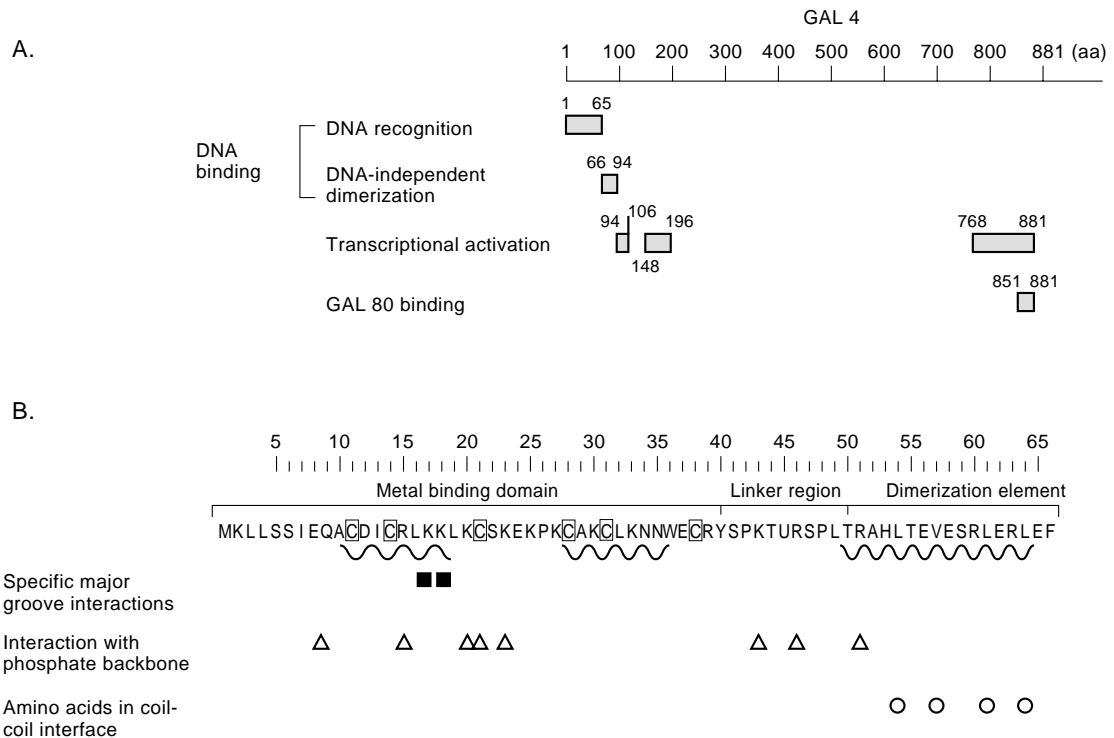


FIGURE 13.14. GAL4 and its DNA-binding site. (A) Domain organization of GAL4. (B) Crystal structure features in graphic form. (Modified, by permission from *Nature* [Marmorstein et al. 1992] Copyright 1992 Macmillan Magazine Ltd.)

Figure 13.15A summarizes chemical and nuclease footprinting results of GAL4 on DNA (Carey et al. 1989). The dyad site and symmetric nature of the GAL4 contacts shown in the figure were the original indication that GAL4 bound to its 17-bp site as a dimer, with each monomer contacting one-half of the dyad. However, when the interactions inferred from DMS, hydroxyl radical, and ethylation interference were displayed on a typical B-DNA helix, as illustrated in Figure 13.15B, left panels (two rotational views are shown), they revealed important information, which led to a more sophisticated model for how GAL4 bound its site. Most of the major groove DMS protections/interference were at G residues located at both ends of the dyad site (gray G residues), whereas the minor groove/backbone interactions extended from the Gs toward the center of the site. The affected phosphates implicated by ethylation interference are in black along the backbone and the affected sugar residues implicated by hydroxyl radical are indicated in boldface stick representations. On the basis of the heterodimer data shown in Figure 13.4 and chemical data in Fig. 13.15A, it was proposed in 1989 (Carey et al. 1989) that each monomer of GAL4 “makes sequence specific contacts in the major groove at the outer base pairs of the recognition site and then snakes along one strand of the phosphate backbone” with the “dimer contacts positioned over the center of the dyad.” Remarkably, the structure solved in 1992 (Marmorstein et al. 1992) confirmed and significantly refined this simple model. The structure revealed that the interactions identified by chemical probing very closely match those interactions identified in the crystal structure.

This point is best illustrated by superimposing the crystal structure of GAL4 onto the DNA and indicating the contacts inferred from chemical probing. Again, two rotational views are shown in the right panels of Figure 13.15B. First, it is clear that GAL4 in the

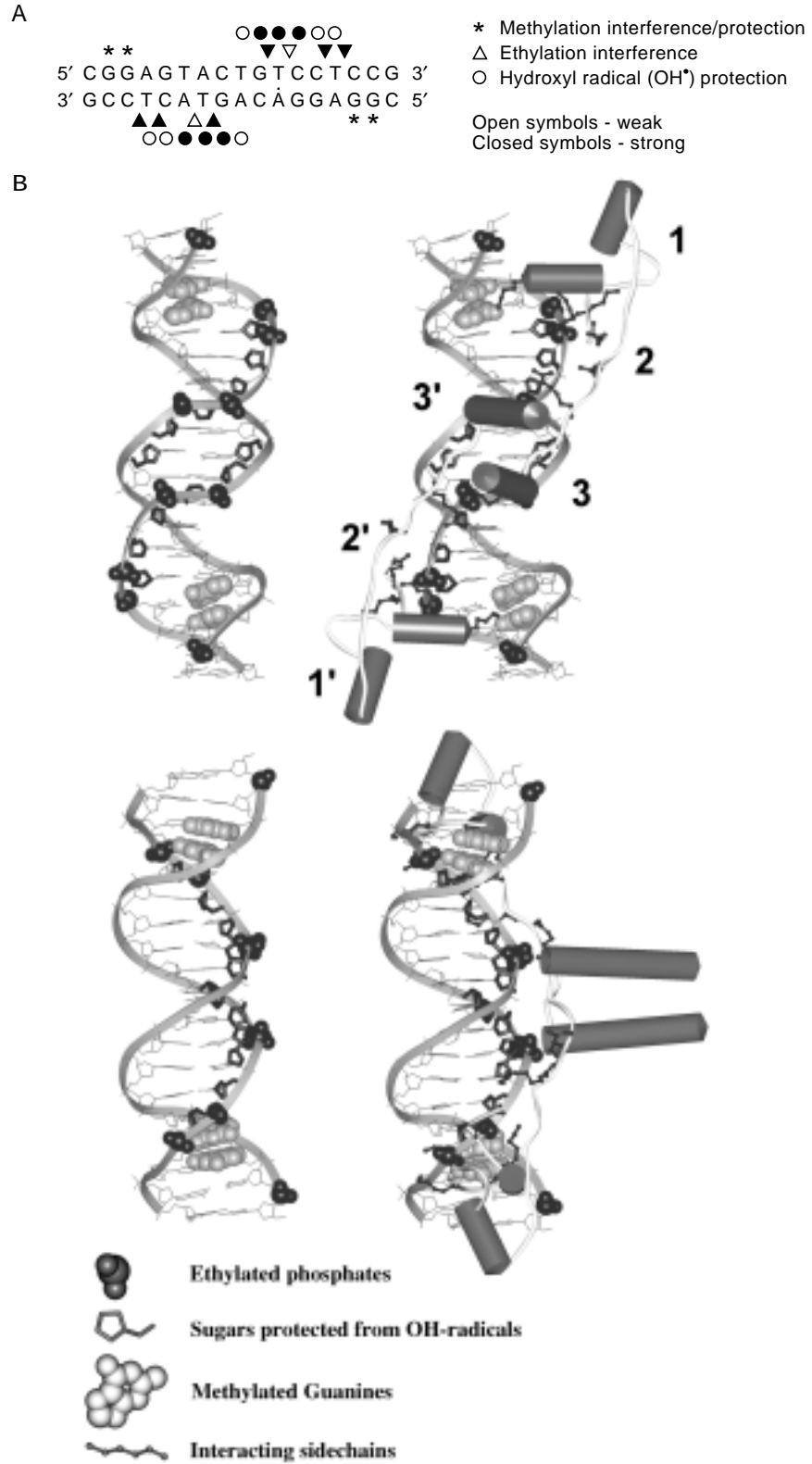


FIGURE 13.15. (A) Chemical contact summary of GAL4 interaction with DNA. (Modified, with permission, from Carey et al. 1989 [Copyright Academic Press Ltd.].) (B) Crystal structure of interaction. (Crystal structures rendered by Michael Haykinson [UCLA] using the Molecular Graphics structure modeling computer program Insight II.)

structure interacts directly with the residues on DNA implicated in contact by chemical probing. More specifically, discrete structural features of the protein generate the contacts with DNA. The zinc-complexed DNA recognition module is located between amino acids 8 and 40 (see Fig. 13.14B). In this domain, the two atoms of zinc from each monomer complex with 6 cysteines to fold into a structure containing two perpendicularly disposed α -helices (Fig. 13.15B features 1 and 1'). One of the helices from each monomer inserts into the major groove perpendicular to the DNA axis and makes sequence-specific contacts with lysines 17 and 18 (black stick representations emerging from the cylindrical helix) and the exposed groups of the G residues at the end of the dyad (indicated as gray van der Waals representations). These Gs are those detected by DMS protection and interference in chemical probing experiments (Fig. 13.15A). Within the superfamily of fungal binuclear cluster transcription factors, most members contain conserved lysines or arginine at position 17 and lysine or histidine at position 18. For many of the known proteins, the contacts of these residues with a dyad CGG motif are a conserved feature of the mechanism of DNA recognition (Liang et al. 1996; Schwabe and Rhodes 1997; Swaminathan et al. 1997). Thus, the GAL4-DNA chemical probing data and the structure can be applied to this very large family of factors to help determine how they recognize their sites.

Immediately carboxy-terminal to the DNA-binding module in each monomer is a flexible unstructured linker (2 and 2') that traces along the minor groove. Mainly lysines and arginines in this linker engage (see Fig. 13.15B) in nonspecific contacts with the phosphate backbone until it comes to the center of the dyad. The black stick representation shows the side chains projecting onto the DNA. These side chains interact with the residues implicated in ethylation interference and hydroxyl-radical protection (phosphates are in black van der Waals, and sugars are shown as sticks). The protein linker varies in length among members of the binuclear cluster family. The variation allows the different members to bind sites with different distances separating the 5' and 3' CGG motifs (Schwabe and Rhodes 1997).

The two monomers each form an α -helix in the center of the dyad (3 and 3'). The α -helices from each monomer associate to form a coiled-coil dimerization motif using leucines and a valine to form the hydrophobic interface (Fig. 13.15B). The base of this coiled-coil motif also engages in nonspecific contacts with the minor groove in the center of the dyad site. Again, the indicated ethylation interference and hydroxyl-radical contacts match those residues that interact with the base of the coiled coil in the crystal structure.

In conclusion, the model predicted from the chemical probing data (Carey et al. 1989) is in close agreement with the crystal structure (Marmorstein et al. 1992).

TBP. The TATA box binding protein, or TBP, is an example of a protein that binds exclusively in the minor groove and makes contacts with the base pairs and with the phosphate backbone. DNA sequence comparisons between the different species reveal a relatively conserved TBP core sequence of 180 amino acids required for TATA box binding and interaction with the general transcription factors. This core contains two 80-amino-acid direct repeats. The amino terminus of the TBP gene differs significantly among species, and it has been proposed that this region mediates species-specific interactions with components of the transcriptional machinery required for activation (Burley and Roeder 1996).

Prior to solving the crystal structure of TBP bound to a TATA box, researchers performed extensive chemical footprinting analysis to understand how TBP and TFIID docked with a TATA box. These analyses led to a model whose main elements were later confirmed by the crystal structure. Figure 13.16A illustrates the results of hydroxyl-radical footprinting, ethylation interference, and DMS protection and interference assays. Note the lack of major groove interactions (i.e., N7-methyl G residues) and the preponderance of putative minor groove contacts (i.e., N3-methyl A residues and ethylated phosphates) inferred from the

chemical probes. These data were originally interpreted as support for a model whereby TBP makes primary contact with the minor groove (Lee et al. 1991; Starr and Hawley 1991).

The model was confirmed and extended by the crystal structure of the TBP:TATA co-complex solved independently by the laboratories of Burley and Sigler. Sigler and colleagues solved a yeast TBP co-crystal (Kim et al. 1993b), and the structure solved by Burley and colleagues was that of the *Arabidopsis* TBP (Kim et al. 1993a). As in the previous figure, the chemical contacts are superimposed on the DNA in the left panels (two rotational views) while the TBP structure is shown with DNA in the right panels of Figure 13.16B. The structure contains a 10-stranded antiparallel β -sheet (arrows) in a saddle-like conformation, with four α -helices (cylinders) crowning the upper surface of the saddle. The protein consists of two subdomains (Fig. 13.16B, 1 and 1') corresponding to each direct repeat in the primary sequence. Each subdomain is organized into a sheet 1–helix 1–sheet 2–turn–sheet 3–sheet 4–sheet 5–helix 2 configuration (Fig. 12.7; see Chapter 12). The β -strands constituting the underside or concave surface of the saddle form a large hydrophobic interface that interacts through van der Waals and hydrophobic contacts with the bases. A nominal number of specific hydrogen bonds (Sigler and colleagues identified 6 and Burley and colleagues identified 4) are formed within the 8-bp region in the minor groove. Binding induces a significant DNA deformation causing an 80° bend and unwinds the DNA 110°, resulting in the characteristic bent shape indicated in the figure. The deformation is caused by intercalation of two phenylalanines between the bases at both ends of the TATAAA. This intercalation buckles the bases, causing the overall site to bend. This can be visualized by focusing on the two upper A residues in the DNA, shown in gray van der Waals representation. The figure with TBP docked shows the intercalating phenylalanine inserting between the two As, causing them to buckle. The resulting bend and untwisting maximize the hydrophobic interface between the protein and the minor groove. Salt bridges between basic residues and the phosphates, water-mediated H-bonds, and van der Waals contacts with the sugars are the driving forces behind the interaction. Most importantly, Figure 13.16B shows that when the structure is docked with the DNA, the crystal interactions explain the hydroxyl-radical protections and almost all of the ethyl phosphates that interfere with TBP binding.

In conclusion, we have illustrated two examples, where the details of the chemical probing experiments allowed investigators to propose models for how the proteins interacted with DNA. Once the structures were solved, they confirmed and extended fundamental aspects of the original model. As more structures are solved, the interrelationship between the chemistry and the mechanism of DNA binding should provide ample information to allow an investigator to extrapolate the data to understand how DNA binding is mediated by unknown proteins or proteins that fit into defined families, one of whose members is of known structure.

Analysis of Promoter-specific Multicomponent Nucleoprotein Complexes

At the outset of this chapter, we described how clusters of proteins bound to promoters and enhancers (i.e., enhanceosomes) are probably best described as networks of protein–protein and protein–DNA interactions (see Chapter 1 and Carey 1998). The two driving forces behind the formation of enhanceosome complexes are cooperativity and DNA looping or bending. Cooperativity is a phenomenon whereby proteins assist one another binding to DNA so that together they bind more tightly than each one alone. Put another way, in the presence of its partner, a protein exhibits a lower K_d or higher affinity for its DNA site. DNA bending allows distal interactions to occur. DNA bending occurs when two distal proteins interact with the intervening DNA looping out. Within the persistence length of DNA (~140

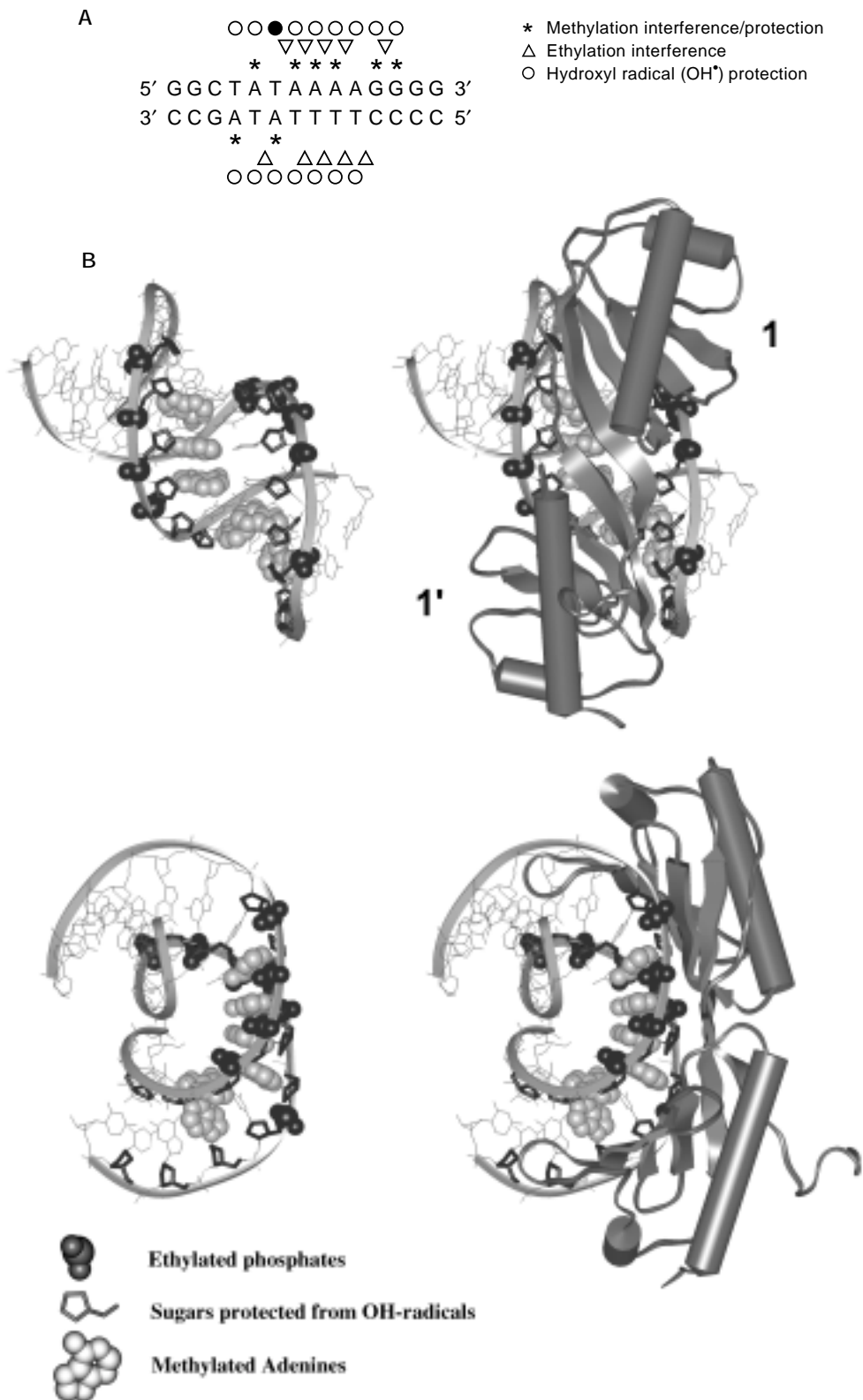


FIGURE 13.16. (A) Chemical contact summary of TBP–DNA interactions. (B) Crystal structure. (Rendered by Michael Haykinson [UCLA] using the Molecular Graphics structure modeling computer program Insight II.)

bp), DNA bending extracts an energetic penalty that must be paid by the strength of the protein–protein interaction or by a protein that stabilizes the bend. Beyond the persistence length, the energetic penalty is less substantial, but the probability of the protein–protein interaction decreases with distance. These effects have been modeled by biophysical chemists but are only beginning to be applied to transcription systems in a predictive fashion (Wang and Giaever 1988; Rippe et al. 1995).

DNA Binding Cooperativity

One of the key mechanisms for generating specificity is cooperative DNA binding. In light of the enhanceosome, this concept is being revisited in the eukaryotic gene expression field. We first discuss the theory and an example followed by a simple approach to studying the phenomenon. Of the many examples of cooperative binding by DNA-bound protein, the paradigm is λ repressor or cI (Johnson et al. 1981). Repressor binding has been analyzed using standard biochemistry and biophysical techniques, and the equilibrium has been analyzed mathematically and thus provides an excellent model for understanding and studying cooperativity (for review, see Hochschild 1991). In a bacteriophage lysogen, repressor maintains the lysogenic state by controlling transcription from two related rightward and leftward operators called O_R and O_L . For simplicity we describe the situation at O_R . At O_R , repressor dimers bind cooperatively to a high-affinity site called O_{R1} and a low-affinity site called O_{R2} . The cooperativity is mediated by direct protein–protein interactions between the carboxy-terminal domains of the adjacent dimers. Repressor bound at O_{R1} represses transcription from a promoter called P_R (promoter in the rightward direction), which activates the lytic cycle. Concurrently, repressor bound at O_{R2} touches its target, RNA polymerase, and activates expression from P_{RM} (promoter for repressor maintenance). P_{RM} controls expression of the λ repressor and is a classic example of an autoregulated promoter.

The cooperative effect was initially measured by DNase I footprinting assays. By comparing the affinity of repressor for a ^{32}P -end-labeled DNA fragment containing O_{R1} and O_{R2} in their natural positions, or either O_{R1} or O_{R2} alone, Ptashne and colleagues showed that the repressor binds cooperatively. As described above, the DNA fragment was set at a concentration below the K_d (10^{-10} M). Then, the repressor was titrated in twofold steps until the DNA fragment was saturated as measured by the footprinting analysis. Recall that if the DNA concentration is below the K_d , then when $PS=S$ (see Box 13.1), conditions where the DNA site is 50% occupied, the $K_d=P$. Put simply, the K_d is equivalent to the concentration of free protein required to generate 50% occupancy of the site. Using this approach, it was shown that the K_d values for the isolated O_{R1} and O_{R2} are 3.3×10^{-9} and 5×10^{-8} , respectively (Fig. 13.17A). However, in the presence of O_{R1} , the K_d for O_{R2} decreases to approximately 10^{-9} . The amount of cooperativity can be expressed as the ratios of the O_{R2} K_d measured in the presence and absence of O_{R1} .

Therefore, in evaluating whether proteins bind cooperatively, their affinities must be measured in isolation and together. In cases where different proteins are binding cooperatively, the same DNA template can be used to perform the analysis. However, in cases where a single protein is binding to multiple sites, the binding of protein to the wild-type template must be compared with templates where one or the other site is mutated. Generally, the cooperativity is easily visualized. Some investigators employ EMSA. In such cases, one protein will enhance the amount of probe shifted by the other in a greater than additive fashion. Often the data can be quantitated using a Hill plot of the bound and unbound DNAs. Although we do not discuss quantitative modeling of the phenomenon, such issues are covered in detailed reviews (Brenowitz et al. 1986; Hochschild 1991).

In a multicomponent complex, all protein–protein interactions should, in principle, have reciprocal effects on each other’s stability. Obviously, in some cases, these effects will be dramatic, such as two proteins binding to weak sites and, in some cases, these will be less dramatic, such as the threefold effect λ repressor bound at O_R2 has on binding to O_R1 (Johnson et al. 1981).

In considering how an enhanceosome might work, one can predict that many of the interactions will be cooperative. As described in Chapter 1, this is indeed the case on both the T-cell receptor α (TCR- α) and interferon β (IFN- β) enhanceosomes (Giese et al. 1995; Thanos and Maniatis 1995). As an extension of this idea, it could be imagined that as the promoter-bound activators help to assemble the general machinery into a transcription complex, the complex should, in turn, help the activators to bind to their sites on DNA. This appears to be the case in the IFN- β enhanceosome (Kim and Maniatis 1997).

DNA Looping and Bending

DNA looping. The ability of two proteins to interact while both are bound to DNA forms the theoretical basis for DNA looping, long believed to be the mechanism by which distally bound activators interact with the general transcriptional machinery. The concept of DNA looping was first discussed in studies on the arabinose operon in *E. coli* and has been implicated in numerous biological processes. The phenomenon has been extensively investigated using the λ repressor system, and we discuss this as an example.

λ repressor as discussed above normally binds cooperatively to two 17-bp sites with a center-to-center distance (inclusive) of 25 bp or 2.4 helical turns. Insertion of 10 or 11 bp (called a helical increment because the periodicity of DNA is 10.5 bp per turn) has little effect on the cooperativity up to 8 turns of DNA away. However, insertion of nonhelical increments between sites separated by up to 8 turns away abolishes cooperativity. Figure 13.17 illustrates an example where insertion of 3.5 helical turns between the repressor binding sites prevents cooperativity, and where insertion of 4 helical turns restores it (Hochschild and Ptashne 1986). Electron microscopy and physical measurements have confirmed that the intervening DNA loops (Griffith et al. 1986). The looping within a plane requires only that the intervening DNA bend smoothly so that the adjacent proteins can interact. The inability of the proteins to interact with the introduction of nonintegral turns is due to the energetic penalty of twisting the DNA. At the distance stated, the cost would be 7 kcal or more. Presumably, the energy of the protein–protein interaction between adjacent dimers is not sufficiently large to maintain the interaction and absorb the cost of twisting. This assertion is supported by the observation that introduction of a short single-stranded gap between the sites, a gap that in principle allows free rotation around the single strand, restores the cooperativity even at nonintegral distances. Surprisingly, cooperativity can be observed up to 20 helical turns away, but the requirement for helical periodicity occurs only up to 8 turns. Presumably after 8 turns the energetic penalty for the twisting is less and the interaction is no longer proscribed (see Rippe et al. 1995 for a recent review on the energetics of looping and twisting). In analyzing a situation where two or more proteins are believed to be interacting, the effect of altering the helical phasing is generally used as evidence for an interaction.

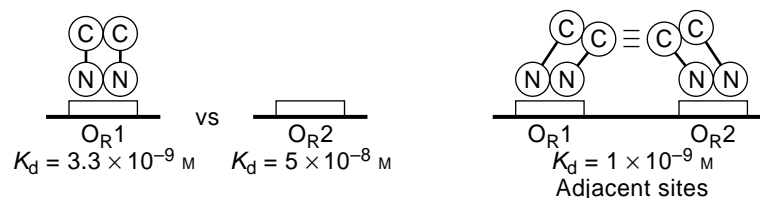
Although looping can be inferred from protein–protein interactions that occur at a distance and generate cooperative binding of a protein to DNA, it is difficult to visualize looping directly. However, the DNase I cleavage pattern of the DNA between the sites can be informative. In the case of repressor bound to sites separated by 6 turns, the DNA between

the sites exhibits a series of enhancements and protections with a 10-bp periodicity (Hochschild and Ptashne 1986). The explanation is that the DNA is smoothly bent, exposing the minor groove on the outside convex surface of the bend and protecting the minor groove on the inside concave surface. Such a pattern might be diagnostic of looping. Long-range interactions between activators in enhanceosomes might also display such periodicity.

There are several features of the λ repressor model system that will likely differ in other systems. Some proteins may be large and sufficiently flexible to interact without an energetic penalty, irrespective of their periodical relationship. Alternatively, one could imagine that the protein-protein interaction is strong enough to absorb the energetic cost. λ repressor indeed represents a novel protein in the sense that its protein-protein interaction, although unable to pay the cost of DNA twisting to bring the proteins into phase, can nevertheless absorb the penalty of in-phase DNA bending. There are many situations where this is not the case, including interactions among activators in the TCR- α and IFN- β enhanceosomes. These interactions require the action of sequence-specific DNA-bending proteins (Grosschedl 1995; Carey 1998).

DNA bending. Studies in prokaryotic and eukaryotic systems over the past several years have emphasized the important role DNA bending plays in DNA metabolic processes. In transcription, the role of DNA bending appears to be to align closely situated proteins into complexes that contact the general transcription machinery. Passive DNA bending within the persistence length of DNA (~140 bp) is constrained due to the rigidity of the DNA helix. Although certain protein-protein interactions can occur by passive bending or looping of the intervening DNA, these require a threshold amount of energy imparted by the protein-protein interaction. If this energy is not sufficient, the protein may only interact with the help of a third protein that has the ability to bend the DNA in a certain direction and to a specific angle, permitting the interaction. These bending proteins are called architectural proteins (Werner and Burley 1997).

A Cooperative binding



B Cooperative binding and DNA looping

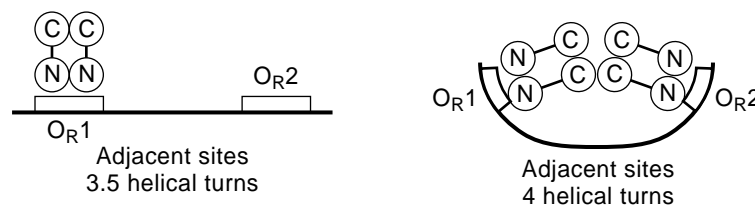


FIGURE 13.17. Cooperative DNA binding and DNA looping.

A prototypic example of a complex multiprotein structure involving cooperative protein–protein interactions, DNA bending, and looping is the intasome. The intasome and other prokaryotic recombination complexes are reviewed in Johnson (1995). The intasome is a protein–DNA complex involved in integrating and excising the 40-kbp bacteriophage λ genome to and from the *E. coli* chromosome. The excision reaction requires the phage-encoded integrase (int) and excision protein (Xis), and the *E. coli*-encoded proteins integration host factor (IHF) and Fis. Both IHF and Fis bend the DNA. The prophage genome is flanked by two attachment (att) sites called attL and attR (Fig. 13.18). Each attachment site contains phage and *E. coli* components; attL is composed of attP' and attB, whereas attR is attP and attB'. All four proteins form a complex on each attachment site on supercoiled DNA (Fig. 13.18), the two attachment complexes interact to form the synaptic complex, and the integrase protein performs a concerted strand exchange and ligation reaction that circularizes the phage genome, excising it from *E. coli* and simultaneously religating the two free *E. coli* chromosome ends.

Within the complex, there are numerous examples of cooperativity between members of the complex, as illustrated in Figure 13.18. Integrase has two DNA-binding domains — one located on the carboxyl terminus, which binds the core sites flanking the recombination crossover point, and one on the amino terminus, which contacts the arms sites. The role of IHF is to bend the DNA to allow the amino- and carboxy-terminal domains to bind their respective sites cooperatively. Xis and Fis promote cooperative binding of Int molecules to P2. The role of IHF is strictly architectural. Thus, although IHF, a sequence-specific heterodimeric protein, can promote cooperative interactions by facilitating DNA looping, it can be replaced by the nonspecific bacterial HU protein or the eukaryotic HMG proteins at appropriate concentrations. It is envisaged that the sole role of IHF is to impart a directional bend. It does not directly contact other proteins or, if it does, the interaction is not essential. The intasome exemplifies how both DNA bending and cooperativity can function to assemble a complex nucleoprotein structure. The methods and approaches used for understanding intasome formation have served as a guide for understanding enhanceosome formation in higher eukaryotes.

IHF, in addition to its role in recombination, participates in gene activation on numerous bacterial promoters. One well-characterized example is activation of transcription by the *Klebsiella* NIF A activator protein, which binds upstream of the σ_{54} RNA polymerase holoenzyme and stimulates its transcription from responsive promoters (Hoover et al. 1990). IHF binding between the NIF A binding site and the nif H core promoter stimulates transcription presumably by bending the DNA and facilitating the ability of NIF A to contact the holoenzyme and induce open complex formation. EM and footprinting analyses have lent credence to this hypothesis.

Mechanisms of DNA Bending

In eukaryotes, DNA-bending proteins bind DNA either specifically or nonspecifically (see Werner and Burley 1997). Examples of specific proteins include: (1) TBP, discussed above, (2) LEF-1 involved in formation of the TCR- α enhanceosome, (3) the sex-determining factor SRY, and (4) HMG(I)Y, which is involved in formation of the IFN- β enhanceosome. LEF-1 contains a domain referred to as the HMG box. This domain was first identified in the eukaryotic HMG 1 and 2 proteins, which are good examples of nonspecific DNA-bending proteins. HMG box proteins bind to the minor groove. Crystal and NMR structures of the HMG box demonstrate an L-shaped structure with three α -helices (Fig. 13.19).

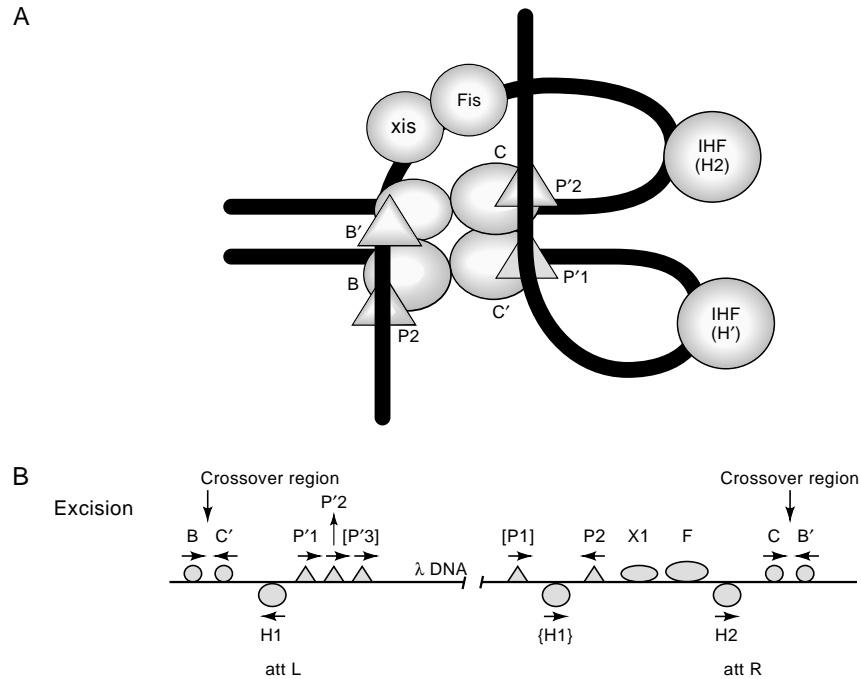


FIGURE 13.18. (A) Model of the intasome. (Redrawn, with permission, from Kim et al. 1990 [Copyright 1990 Cell Press] and Kim and Landy 1992 [Copyright 1992 American Association for the Advancement of Science].) (B) Binding sites for various proteins. (Adapted, with permission, from Johnson 1995 [Copyright Oxford University Press].)

Bending is facilitated by insertion of a more hydrophobic amino acid between the base pairs, resulting in unstacking of the bases, partial unwinding of the helix, and a bend toward the major groove.

Not all proteins that bend DNA bind in the minor groove. The best-characterized of these are prokaryotic regulatory proteins CAP and Fis (Pan et al. 1996). These dimeric proteins bend DNA by inserting recognition α -helices into adjacent major grooves. In the case of Fis, the center-to-center distance of the recognition helices is less than the center-to-center distance of the major grooves, and the protein must bend the DNA to bring the DNA's sites in register with the protein. In such cases the DNA may also wrap more extensively around the protein, further exaggerating the bend. Some eukaryotic proteins, including members of the bZIP family, also bend the DNA, apparently from the major groove.

Approaches for Studying Bending

There are three commonly used methods to measure DNA bending: EMSA, DNA cyclization, and electron microscopy (for reviews on theory and methodology, see Crothers et al. 1991, 1992). EMSA is one of the most tractable and informative methods, as it can provide information on the position, magnitude, and direction of the bend when compared with known standards.

When DNA migrates through a polyacrylamide or agarose gel, it is thought to slither through the gel pores. Although molecules of different length have the same charge density (due to the uniformity of the sugar/phosphate backbone) and, hence, experience the same electrical force, the smaller ones are able to move efficiently through the pores. Bent DNA slows the movement, with the effect being more pronounced when the bend is posi-

tioned in the center of a DNA molecule as opposed to the ends. The current view is that the mobility is inversely proportional to the square of the distance between the DNA ends. In a DNA molecule bent in the center, the distance between the ends is at a minimum and will therefore retard mobility to the greatest extent versus a linear molecule or a molecule bearing a bend positioned near the end of the molecule.

In a typical assay (see Fig. 13.20) the DNA site of interest is cloned into a vector containing two tandem polylinkers separated by a unique cloning site (see Zwieb and Adhya 1994). The DNA molecule is then cut with restriction endonucleases, which cleave once in each of the polylinkers. This generates a series of circularly permuted fragments, identical in sequence composition but differing in the position of the site relative to the ends of the polylinker. The DNA molecules are then labeled with ^{32}P and incubated with protein, and the complexes are fractionated on native polyacrylamide gels. The protein retards all of the fragments relative to unbound DNA. However, if the protein significantly bends the DNA helix, it will retard fragments bearing centrally positioned sites to a greater extent than fragments bearing a distally positioned site. The mobility of the fragments (vertical axis) measured as the distance from the end of the gel is then plotted against the position of the site in the DNA fragment. The curves display a minimum mobility at the bend center.

To determine the directionality of the bend, the binding site is placed on a DNA fragment bearing a phased A tract—stretches of 5 or more consecutive A residues that cause the DNA to bend in a known direction. A linker of various lengths is placed between the A tracts and protein-binding site. As the linker length is expanded or contracted, the protein-induced bend and the A-tract bend are rotated relative to one another. Under such condi-

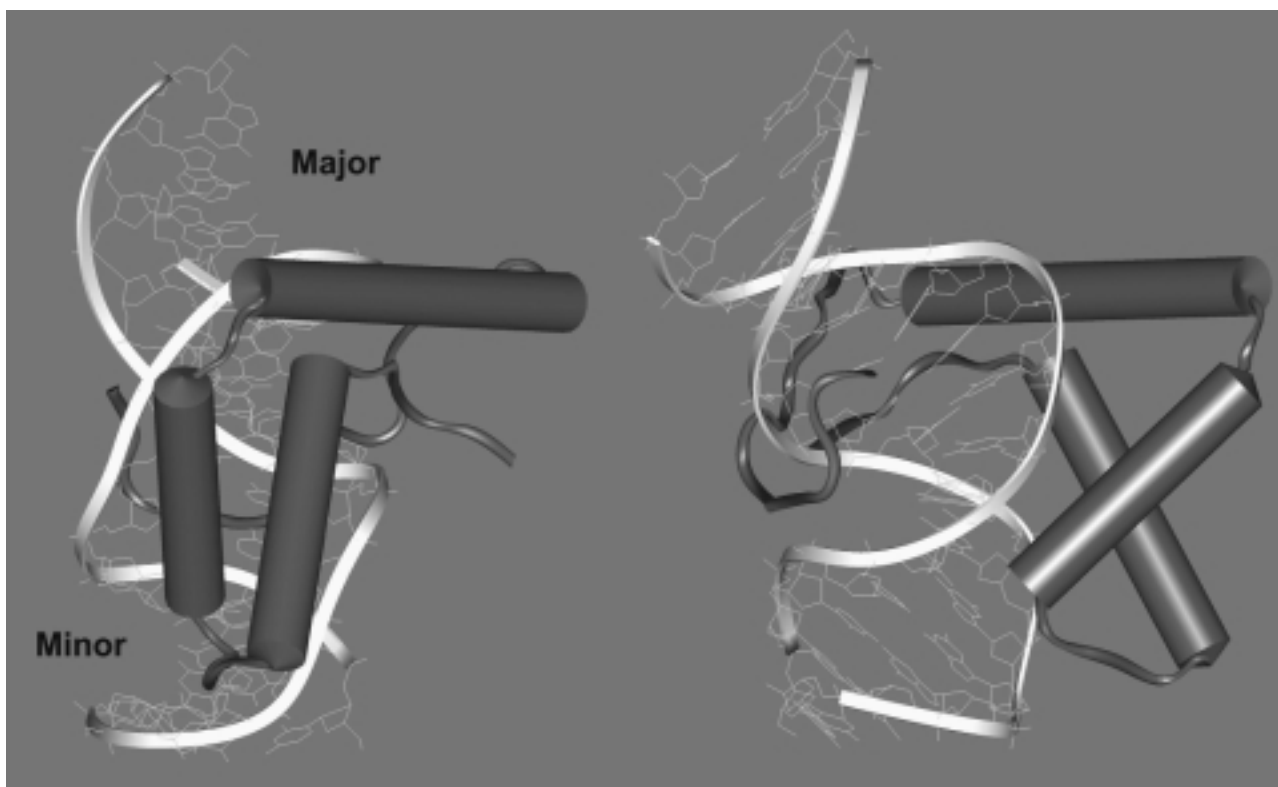


FIGURE 13.19. LEF-1 binding in minor groove and resulting DNA bend. (Crystal structures rendered by Michael Haykinson [UCLA] using the Molecular Graphics structure modeling program Insight II.)

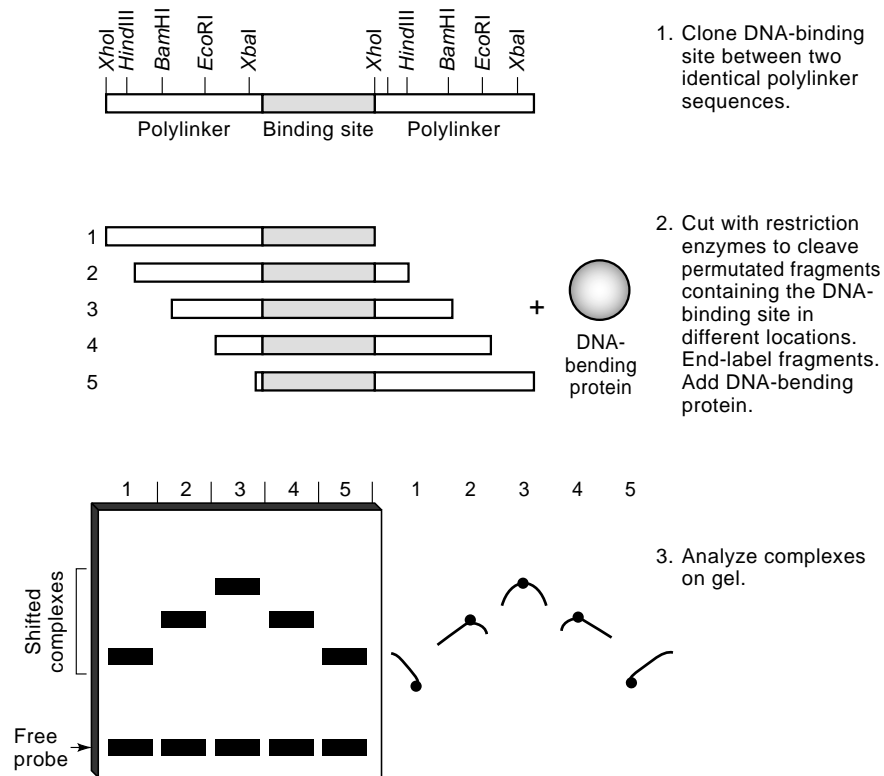


FIGURE 13.20. Analysis of DNA bending.

tions, the fragment will display a sinusoidal mobility pattern. When the bends are in the same direction, the mobility of the fragment reaches a minimum. At this point, one can be confident that the bend induced by the protein is in the same direction as the known bend induced by the A tracts. This bend can be assessed in the context of its natural site for its biological relevance.

The angle of A-tract bending has also been rigorously established and can therefore be used as a standard to determine the bend induced by the protein. In this experiment, the protein-binding site is again placed at the end of a DNA molecule containing phased A tracts. As the number of A tracts increases, so does the bend angle. Because the A tracts are separated by 10–11 bp, the direction of the bend for each A tract will be identical, and as additional A tracts are added, the bends add. The presence of a protein-binding site on the end allows for correction of the mobility. The protein–DNA complex with the bend site placed in the center is now compared with the A-tract standards (Crothers et al. 1991).

In summary, a concerted study of how eukaryotic activators bind DNA and assemble into enhanceosomes requires a significant effort to apply numerous methodologies to a highly sophisticated problem. The study of such complexes in the prokaryotic field took more than a decade and the commitment required was great. Nevertheless, the assembly and regulation of such nucleoprotein complexes almost certainly hold the key to the phenomenon of specificity during combinatorial control and gene regulation.

TECHNIQUES

PROTOCOL 13.1

DNase I Footprinting

The DNase I footprinting protocol was introduced to the research community in 1978 (Galas and Schmitz 1978). The method was based on the nested-set theory that formed the conceptual basis for DNA sequencing (Maxam and Gilbert 1977). Because of its simplicity, DNase I footprinting has found a wide following for both identifying and characterizing DNA–protein interactions.

The concept is that a partial digestion by DNase I of a uniquely ^{32}P -end-labeled fragment will generate a ladder of fragments, whose mobilities on a denaturing acrylamide gel and whose positions in a subsequent autoradiograph will represent the distance from the end label to the points of cleavage. Bound protein prevents binding of DNase I in and around its binding site and thus generates a “footprint” in the cleavage ladder (see Box 13.3 and Fig. 13.8). The distance from the end label to the edges of the footprint represents the position of the protein-binding site on the DNA fragment. The exact position of the site can be determined by electrophoresing a DNA sequencing ladder alongside the footprint.

DNase I cannot bind directly adjacent to a DNA-bound protein because of steric hindrance. Hence the footprint gives a broad indication of the binding site, generally 8–10 bp larger than the site itself. Furthermore, the crystal structure and extensive biochemical studies on DNase I (for review, see Suck 1994) show that it binds in the minor groove, contacts both strands of the sugar/phosphate backbone, and bends the DNA toward the major groove. The sequence-dependent width of the minor groove at a particular position will influence the cleavage efficiency. As a result, the cleavage ladders will not be uniform and certain gaps will exist. In some cases, proteins bind in these gaps and the footprints on such sites are not as dramatic as when the binding site is positioned over a region where the cleavage is more uniform and efficient. Footprinting can be performed with purified proteins or with crude extracts. The technical considerations for pure and crude systems are elaborated upon below and in Chapter 8.

TIME LINE AND ORGANIZATION

A protocol for performing a DNase I footprinting analysis is described below. Generally, one must prepare a uniquely end-labeled DNA fragment encompassing the protein-binding site. For small sites, double-stranded oligonucleotides that are labeled on a single strand are satisfactory. A highly active stock of DNase I must be obtained and prepared from a powder or obtained in a solution form. An entire analysis generally takes 3 days, because first the optimal concentrations of DNase must be determined before performing a dose–response curve with the protein of interest. Ideally the investigator has either a pure protein or an extract expressing the protein and preliminary data from mutagenesis studies, transfections, or EMSA indicating that a protein does indeed bind to the site.

Day 1: Prepare ^{32}P -end-labeled fragment (described in Protocol 13.6)

Day 2: Perform DNase I titration to optimize amounts and conditions for footprinting

Day 3: Use optimal amounts of DNase I to perform a dose–response curve with protein of interest

OUTLINE

DNase I titration (time commitment: 1 day)

Step 1: Pour gel from premade acrylamide/urea mix (30 minutes)

Step 2: Prepare buffers for DNase I footprinting (1 hour)

Step 3: Reactions: Prepare reactions and titrations (30 minutes), and incubate reaction mixtures at appropriate temperature (15 minutes)

Step 4: DNase I footprinting: Perform DNase I dilutions (15 minutes). Add DNase I and perform footprinting reactions (5 minutes). Terminate reactions and process products (1 hour)

Step 5: Gel electrophoresis (4–6 hours) and autoradiography (8–12 hours) or phosphorimager analysis

Step 6: Data analysis

Dose response curve (1 day)

Steps 1 and 2: As above

Step 3: Prepare serial dilutions of protein (30 minutes)

Step 4: DNase I footprinting as above

Step 5: As above

PROCEDURES

CAUTIONS: *Acrylamide, CaCl₂, Chloroform, DTT, Ethanol, Formamide, Glycerol, KCl, KOH, β-Mercaptoethanol, MgCl₂, Phenol, PMSF, Radioactive substances, SDS. See Appendix I.*

*DNase I TITRATION***Step 1: Pour gel from premade acrylamide/urea mix**

About 2 hours before beginning the experiment, pour an 8–12% polyacrylamide/urea gel depending on the fragment size. The gel will take about 1 hour to polymerize and 30 minutes to 1 hour to pre-run prior to sample loading. See Sambrook et al. (1989, pp. 13.45–13.57) for instructions on mixing and pouring acrylamide gels.

Step 2: Prepare buffers for DNase I footprinting

Four buffers are required for the DNase I footprinting protocol in addition to the protein. The amounts needed will have to be determined empirically depending on the scope of the study.

Buffer D: We often use as a DNA-binding buffer a buffer referred to as Buffer D from the Dignam and Roeder protocol for preparation of nuclear extract (Dignam et al. 1983). Nuclear extract is dialyzed against Buffer D containing PMSF and DTT. However, any buffer of similar composition is compatible with the assay. The key point is that the final salt and magnesium concentrations must be compatible with protein binding in vitro and with the catalytic activity of the DNase I. Buffer D alone has a shelf life of several months at room temperature but should ideally be filter-sterilized before long-term storage.

Buffer D (0.1 M KCl):

20 mM HEPES-KOH (pH 7.9)

20% glycerol

0.2 mM EDTA

0.1 M KCl

Protein diluent buffer: This is the buffer into which the DNA-binding protein is diluted. However, in crude extracts the protein is not diluted. In this model case the diluent is Buffer D containing a protein-stabilizing agent, such as nuclease-free BSA, an anti-aggregation agent such as 0.01–0.1% NP-40 or Triton X-100, and a reducing agent such as 0.1–1 mM DTT or 10–50 mM β -mercaptoethanol. These reducing agents are particularly important for protein with cysteines within the DNA-binding domain (i.e., zinc fingers or the basic region of many b-Zip proteins). This buffer should be prepared in small quantities and stored in aliquots in a freezer. Dilute solutions of reducing agents generally have short half-lives, so long-term storage of the solutions on ice or in the refrigerator is not recommended.

500 μ l of Buffer D (0.1 M KCl)
 +1 μ l of BSA (50 μ g/ μ l)
 +1 μ l of β -mercaptoethanol (14 M)

DNase I diluent buffer: This buffer is the same as the protein diluent and is designed to be compatible with it. CaCl_2 is added to enhance the activity of DNase I, which is essential when footprinting crude extracts or impure preparations of protein. This buffer should be prepared fresh and kept on ice or at 4°C.

500 μ l of protein diluent buffer (0.1 M KCl)
 +20 μ l of 1 M CaCl_2

Stop buffer: The Stop buffer terminates the DNase I reaction and prepares it for phenol extraction and ethanol precipitation. EDTA is added to chelate divalent cations necessary for DNase I activity; SDS is added to denature protein and strip them from DNA; sodium acetate is added to facilitate ethanol precipitation as is the carrier tRNA. This solution can be maintained as a stock for several months at room temperature.

Final concentrations
 400 mM sodium acetate
 0.2% SDS
 10 mM EDTA
 50 μ g/ml yeast tRNA

Immediately before using, aliquot the amount of Stop buffer needed for the experiment and add 1/1000 volume of Proteinase K solution (10 mg/ml in deionized, distilled water).

Formamide dye mix:

98% deionized formamide
 10 mM EDTA (pH 8)
 0.025% xylene cyanol FF
 0.025% bromophenol blue

Step 3: Reactions

For the typical experiment using a recombinant DNA-binding protein, we initially prepare 12 reactions, which are used to titrate both the initial DNase I and DNA-binding protein concentrations. Higher concentrations of a DNA-binding protein, particularly with crude extracts, tend to inhibit the DNase I. Therefore, the concentration of DNase I used for naked DNA controls will be lower compared with the concentration used with high concentrations of a DNA-binding protein. Furthermore, the amount of DNase I required will

be smaller with increasing fragment size. It is important in this initial step of the study to optimize the concentrations of DNase I needed for a variety of conditions.

1. Prepare a reaction mix containing the reagents described below.

Note: All of the pipetting and reaction preparations are ideally done on ice. Prepare enough mix for two extra reactions, one for inevitable losses or inconsistencies during pipetting and one for a mock reaction without DNase I.

Mix:

DNA template ^a (50 fmole/ μ l)	0.10 μ l (1000–10,000 cpm)
poly dI:dC ^b (1 μ g/ μ l)	0.20 μ l (200 ng)
BSA (50 μ g/ μ l)	0.20 μ l (10 μ g)
β -mercaptoethanol (14 M)	0.10 μ l
MgCl ₂ (0.1 M) ^c	1.5 μ l
Buffer D (0.1 M KCl) ^d	11.5 μ l
H ₂ O	5.4 μ l
<hr/>	
Total volume	19.00 μ l
14 Reaction batch	266 μ l

Recombinant activator protein is prepared at a concentration of 250 ng to 1 mg per ml. For the DNase I titration, 1 ml of neat, undiluted protein and a 1/40 dilution will be used in the initial DNase I titrations. For the dose-response curves (day 3), we perform a titration in threefold steps based on the results from day 2.

^a End-labeled activator-specific DNA probe (Protocol 13.6 for end-labeling DNA).

^b Amersham Pharmacia Biotech Inc. (#27-7880-01)

^c For crude extracts that may contain endogenous nucleases, binding reactions may work better if set up in the absence of MgCl₂, which can be added later in the DNase I diluent buffer.

^d For crude extracts, the Buffer D should be substituted by extract in Buffer D allowing at least 12 μ l to be added. For less concentrated extracts, Buffer D and H₂O can both be substituted allowing 17 μ l of extract to be added.

2. Aliquot the reaction mix into 13 0.5-ml siliconized eppendorf tubes on ice.
3. Thaw recombinant activator protein on ice immediately before titrating. Add 1 μ l to 39 μ l of protein diluent to make the low-concentration sample. Add protein and mix gently by tapping the tube with your finger (vigorous mixing using the pipetman is not recommended, as some proteins may be sensitive to the frothing often generated in this technique). Once the dilution is made, titrate 1 μ l of neat activator protein into the first four tubes, 1 μ l of the 1/40 dilution into the next four tubes, and 1 μ l of protein diluent into the final four:

Tubes 1–4: 1 μ l of neat activator protein

Tubes 5–8: 1 μ l of a 1/40 dilution of activator protein

Tubes 9–12: 1 μ l of diluent alone

For crude extracts, DNase I must be titrated with low and high amounts of extract. Note that crude extracts are highly inhibitory and large amounts of DNase must be added.

4. Place the tubes into a 30°C water bath for 20 minutes (or for crude extracts at room temperature or on ice). This is generally enough time for binding of an activator protein to reach equilibrium, although some proteins (e.g., TBP) bind more slowly.

Step 4: DNase I footprinting

The goal here is to determine roughly the amount of DNase I to employ for a given amount of DNA-binding protein. We therefore make two concentrations (high and low) of protein and, for each, add increasing concentrations of DNase I. We intend to identify concentrations of DNase I that generate an evenly distributed cleavage ladder, where ~50% of the DNA remains uncleaved. This amount of cleavage minimizes multiple cleavages within a single DNA molecule, which can affect the interpretation of data.

1. Choose four concentrations of DNase I for the initial titration.

Note: The concentrations required will be highly dependent on the source of DNase I and the divalent cation concentrations in the mixture as well as the DNA-binding protein being employed. We typically use the Amersham Pharmacia Biotech Inc. FPLC-pure, RNase-free DNase I (cat. # 27-0514-01), which is delivered at a concentration of 5,000–10,000 units per ml.

2. Dilute the DNase I serially to 1/9, 1/27, 1/81, and 1/243 using DNase I diluent. The amount of DNase I added to a cleavage reaction will be 1 μ l. Note that for fragments larger than that used here, smaller amounts of DNase I will be required.
3. After the 20-minute incubation in Step 3.4, perform sequential 1-minute DNase I digestions as follows:

Use timer:

add 1 μ l of DNase I to tube #1 at time n

add 1 μ l of DNase I to tube #2 at time $n+15$ seconds

add 1 μ l of DNase I to tube #3 at time $n+30$ seconds

add 1 μ l of DNase I to tube #4 at time $n+45$ seconds

add 100 μ l of stop buffer + Proteinase K to tube #1 at time $n+1$ minute

add 100 μ l of stop buffer + Proteinase K to tube #2 at time $n+1$ minute, 15 seconds

add 100 μ l of stop buffer + Proteinase K to tube #3 at time $n+1$ minute, 30 seconds

add 100 μ l of stop buffer + Proteinase K to tube #4 at time $n+1$ minute, 45 seconds

Process one set of four reaction tubes per 2 minutes. It is convenient to place the four DNase I dilutions immediately behind the tubes into which you will be titrating. Remember to mix gently by tapping the tubes immediately after pipetting in the DNase I. Place a radioactive disposal vessel nearby, because the pipet tips will be mildly radioactive. For crude extracts, the DNase I digestions, like the binding reactions, may work better if performed on ice.

4. Repeat this entire cycle as soon as possible for the remaining two sets of tubes.
5. Perform a mock reaction alone without any DNase I gel so that in the final analysis the uncut probe can be compared to the cleavage ladders generated above.
6. Incubate all reaction tubes at 55°C for 15 minutes to allow Proteinase K digestion to occur. This is particularly important for cruder protein preparations. Meanwhile, assemble the polymerized gel into the gel apparatus. Pre-run the gel at 1000 V for 30 minutes.
7. Extract each reaction mix with an equal volume (100 μ l) of phenol.
8. Centrifuge tubes for 2 minutes.

9. Transfer a constant volume (e.g., 100 μ l) of the top (aqueous) phase to 1.5-ml non-siliconized eppendorf tubes.
10. Repeat steps 7–9 using an equal volume (100 μ l) of phenol/chloroform.
11. Add 2x volumes of cold 95% ethanol to the 1.5-ml tubes to precipitate DNA. Mix well.
Note: There is no need to add 3 M sodium acetate because the stop buffer contains high salt.
12. Incubate the tubes on dry ice, dry-ice ethanol, or in the -80°C freezer for 10 minutes.
13. Centrifuge the tubes at 14,000g in a microfuge for 10 minutes.
14. Decant ethanol from pellets using drawn-out, narrow-bore, pasteur pipets. The pellets are loose, so be careful!

Note: Monitor tubes periodically with a Geiger counter to ensure that the pellet has not been accidentally drawn up into the pasteur pipet.

Wash the pellets with 100 μ l of 80% ethanol, centrifuge for 2 minutes, and remove the alcohol with drawn-out pipets. Allow the pellets to air-dry completely.

Step 5: Gel electrophoresis and autoradiography

1. Resuspend the pellets in a formamide dye mix.
2. Denature DNA by incubating samples for 2 minutes at 95°C .
3. Load the samples on a polyacrylamide-urea gel and perform gel electrophoresis.
4. Dry the gel under vacuum.
5. Perform autoradiography or phosphorimager analysis.

Step 6: Data analysis

The purpose of the DNase I titration is to optimize cleavage so that an evenly distributed cleavage ladder is obtained and at least 50% of the probe remains uncleaved. To assess the amount of cleavage accurately, the digestion ladders must be compared with a mock reaction containing probe alone but no DNase I. Also, with crude extracts, the mock reaction performed with extract reveals the presence of endogenous nucleases. Ideally, because the DNase I concentration is increased, there should be a concomitant decrease in the amount of intact probe and an increase in the amount of cleavage products. As the probe begins to disappear, there should be a change in the distribution of cleavage products from lower mobility to higher mobility bands as the DNase I begins to cleave the probe into smaller fragments. Ideally, the amount of DNase I that generates approximately 50% cleavage at each concentration of DNA-binding protein should be noted and used to calculate amounts for the dose–response curve.

In the next step, we normally perform a dose–response curve from 1 μ l of DNA-binding protein to a 1/2000 dilution. Although it is unnecessary to adjust the DNase I for every protein concentration, it should be adjusted at concentrations at which the protein is likely to be inhibitory. We adjust the concentration based on the difference in inhibition between 1 μ l of activator protein and the 1/40 dilution. We assume that the inhibition will be somewhat linear with respect to DNase I concentration and extrapolate the amounts accordingly.

PERFORMING A DOSE-RESPONSE CURVE

Step 1: Pour gel from premade acrylamide gel mix (Sambrook et al. 1989, pp. 13.45–13.57)

Step 2: Prepare buffers for DNase I footprinting

Step 3: Prepare serial dilutions of protein

1. Set up the reaction mixtures as described (pp. 473–474) except now add the recombinant activator protein in serial dilutions using the protein diluent buffer. Start with 1 μ l of activator protein and perform threefold serial dilutions for about 8 points (1, 1/3, 1/9, 1/27, 1/81, 1/243, 1/729, 1/2187). For crude extracts, 1 to 17 μ l should be titrated.
2. Prepare a batch for 10 reactions as in step 3.1, aliquot into 9 reaction tubes, and add 1 μ l of the protein dilutions to each tube. Prepare one tube without protein for comparison.
3. Choose the optimal amount of DNase I determined on day 2. Add DNase I to each reaction. Perform digestion and process products as described above.

Step 4: Gel electrophoresis and autoradiography or phosphorimager analysis

Ideally, there should be a gradual increase in the appearance of the footprint as the DNA-binding protein concentration increases. The fractional occupancy or percent protection, two terms useful in affinity calculation, can be deduced by comparing bands from the cleavage ladder in the no-protein lanes with those in the footprint lanes using laser densitometry or phosphorimaging software. An unaffected band below and above the footprint can be used to normalize the DNase I cleavage efficiency. Most proteins will bind a DNA site gradually and follow normal Michaelis kinetics for a first-order reaction. However, many proteins bind as dimers and higher oligomers, and if the dimerization constant is near to the K_d for the site, protein binding may follow second-order or higher kinetics. If this technique is to be used quantitatively to model DNA-binding kinetics, read Koblan et al. (1992) for a more sophisticated overview of technical considerations.

ADDITIONAL CONSIDERATIONS

Generalizing the protocol

There is no steadfast rule for generalizing the protocol. The following factors should be considered:

- Smaller reaction sizes will favor less concentrated DNA binding proteins.
- Lower salt concentrations will favor weaker binding proteins.
- The final volume can be anywhere from 10 to 100 μ l, depending on the reaction.
- Smaller reaction volumes can be used to conserve valuable materials.
- Different carrier DNAs can have different effects: Poly dI:dC, for example, is reported to compete for binding of TBP to a TATA box probe and is often replaced with poly dG:dC.

The following are some ranges that have been used successfully in our laboratories and the published work of others:

- 10–50 mM HEPES-KOH or Tris-HCl (pH 7.0–8.0). DNase I is compatible with a wide range of buffers.
- 0–10 mM MgCl₂. This divalent cation neutralizes the phosphates, and some DNA-binding proteins have been reported to bind it as a co-factor.
- 50–100 mM KCl, NaCl, potassium acetate, ammonium sulfate. Although most DNA-binding proteins are sensitive to unusually high concentrations of salt, salt optima can vary considerably depending on the protein. It is best to evaluate this empirically.
- 0–20% v/v glycerol. Glycerol is a stabilization reagent that lowers the water concentration in the reaction and mimics the in vivo environment. It is also a free-radical scavenger.
- 10–100 µg/ml BSA. BSA is another protein stabilizer, which can act as nonspecific carrier and prevent spurious attachment of dilute protein samples to surfaces (e.g., the walls of eppendorf tubes).
- 0.01–0.1% NP-40, Triton X-100. These nonionic detergents prevent nonspecific protein binding to surfaces but also act as anti-aggregation agents.
- 0.1–1 mM DTT or 10–50 mM β-mercaptoethanol. Absolutely essential reducing agents for many proteins.
- 0–1 µg carrier DNA, either calf thymus or synthetic co-polymers (dI:dC). Prevents nonspecific binding of contaminant during binding reactions and minimizes nonspecific binding of the protein being footprinted.
- Polyvinyl alcohol (PVA), polyethylene glycol (PEG), dimethyl sulfoxide (DMSO). PVA and PEG reagents can vary considerably in concentration. PVA and PEG are volume exclusion agents that increase the functional concentration of the protein in solution and decrease water concentration. DMSO is a denaturant that has some unusual stabilizing properties at low concentration, possibly by minimizing nonspecific binding of the protein or favoring conformational flexibility.

A detailed analysis of relevant parameters and how these influence DNA–protein interactions can be found in reviews by Record et al. 1991 and Koblan et al. 1992.

DNA fragment size

There are several important considerations for fragment size. Highly purified proteins in which the mechanism of DNA binding or affinity is being established should utilize small DNA fragments, such as sites cloned into the polylinkers of vectors such as pGEM or pUC (i.e., 50–100 bp in length). The small size results in better band resolution on polyacrylamide/urea sequencing gels and allows the same fragment to be compared in EMSA and footprinting studies. When attempting to locate a protein-binding site on a promoter, by necessity larger fragments must be employed (see Chapter 8). In crude extracts, where there is an abundance of nonspecific DNA-end-binding proteins like the Ku autoantigen, the sites of interest are ideally located 50 bp or so from the ends, so the footprint can be distinguished from the contaminating end-binders. Please note that for fragments larger than that used here, smaller amounts of DNase will be required.

Cruder extracts

Use of crude extracts in footprinting studies adds an additional complexity to the experimental design. Higher concentrations of DNase I must be employed to overcome the abundance of inhibitors. Some extracts are rich in nonspecific DNA-binding proteins and, consequently, low concentrations of extract (<10 μg) and a high concentration of nonspecific carrier DNA (1 μg or more) must be employed to minimize the inhibition. This must be balanced against the possible need for high concentration of extract to detect specific binding. Due to the presence of endogenous nucleases, it is prudent sometimes to perform the initial binding and DNase I reactions in the cold (4°C). However, higher concentrations of DNase I are necessary to cleave at lower temperatures, although these amounts must be determined empirically. Alternatively, MgCl_2 can be omitted from the preincubation to inhibit nuclease activity during the binding reaction. MgCl_2 is then added with the DNase I.

Time course

It is necessary occasionally to measure the kinetics of binding by a protein. Ideally, this means staggering the time that the protein is added to the binding mixture and the time that the mixture is placed at the proper incubation temperature. Simply adding the protein to a mixture on ice and assuming it will not bind to DNA until the mixture is incubated at 30°C is incorrect. Many binding studies are in fact performed on ice, as alluded to above. Thus, ideally the reaction mixture is brought to incubation temperature and then the DNA-binding protein is added.

Indeed, the rates at which some reactions reach equilibrium are very rapid (i.e., binding of GAL4-VP16 is complete within 2 minutes), whereas others are slow (binding of TBP can take an hour or more). In the former case, it is prudent to speed up the DNase I cleavage reaction considerably by increasing the concentration of DNase I and lowering the time of cleavage to about 15 seconds. Conversely, for slow reactions, the DNase I cleavage rate can be lowered by lowering the concentration of DNase I. Others have commented on the conditions needed for performing truly quantitative footprint for use in biophysical modeling (Brenowitz et al. 1986).

Using unlabeled DNA

It is possible to perform DNase I footprinting on circular DNA templates and to detect the cleavage points by indirect end-labeling or LM-PCR (see Chapter 10, Protocol 10.2 or Protocol 15.2 in Chapter 15). These methods are described in Gralla (1985) and Grange et al. (1997).

TROUBLESHOOTING

No cleavage

Possible cause: Sometimes nuclear extracts can be highly inhibitory at commercially available DNase I concentrations

Solution: If no cleavage by DNase I is observed, increase the DNase I, MgCl_2 , or CaCl_2 concentrations and, most certainly, try a different batch of DNase I. Try lowering the extract concentration or further fractionating it by column chromatography. Additionally, purchase DNase I powder and dissolve it at higher concentrations.

No or weak footprint

Possible cause: This result could be due to low concentrations of protein, the absence of a physiological site in the fragment, or inhibitors of binding contaminating protein preparations.

Solution: Attempt to raise protein concentration, lower DNA concentration, or adjust binding parameters to optimize salt concentrations, pH, etc.

Smeary gels

Possible cause: Smeary gels are most often caused by: (1) not pre-running the gels to remove the residual ammonium persulfate in the acrylamide; (2) incomplete removal of protein during the phenol extraction step and subsequent co-precipitation with alcohol making the resulting pellet difficult to resuspend; or (3) the pellet contains too much salt.

Solution: In case (1), occasionally the bands become compressed at the salt front. Next time, pre-run the gel. In case (2), one observes black spots that smear down the gel from the wells. Try resuspending the sample more vigorously or phenol-extracting the sample one extra time before ethanol precipitation. Another option is to precipitate the sample with 2 M ammonium acetate, which tends to leave protein, small oligonucleotides, and nucleotides in solution while the larger DNA fragments precipitate. In case (3), often a reverse dovetail effect is observed where the banding pattern narrows as it approaches the lower part of the gel. Simply wash the pellet with 80% ethanol before drying and resuspending in formamide dye buffer.

PROTOCOL 13.2

Hydroxyl-radical footprinting

The hydroxyl-radical footprinting methodology was devised by Tullius and Dombroski in 1986 initially to study λ repressor and Cro binding to DNA. However, the method has wide applications to studying protein–DNA interactions, as well as structural perturbations (e.g., bending) that occur in DNA during protein binding. Hydroxyl radicals cleave DNA by abstracting a hydrogen from C4 of the sugar in the minor groove. Protein binding over the minor groove generally protects the sugar from cleavage (see Box 13.4 and Fig. 13.10). The method can give detailed information on protein binding to the minor groove. Although some of the information overlaps that obtained by ethylation interference, it is often complementary and can assist in modeling how a protein docks with its site. As described earlier, the radical is generated by Fe(II) EDTA, which cleaves hydrogen peroxide into a hydroxyl radical and a hydroxide ion. The radical then cleaves the sugar in a diffusion limited reaction. Ascorbic acid is included to regenerate the active Fe(II).

Hydroxyl radical is highly reactive but easily quenched by glycerol; therefore, glycerol, a component in most binding reactions, must be avoided. The radical reaction requires peroxide, which is a powerful oxidant and occasionally interferes with protein binding to DNA. The conditions can occasionally be adjusted to accommodate such scenarios. The radical also reacts with and cleaves proteins; long reaction times can lead to protein degradation.

TIME LINE AND ORGANIZATION

The reaction is set up almost identically to DNase I footprinting except for the absence of glycerol in the buffers. Often, the DNA plasmid quality is essential to the success of the method. Many preparations contain a low level of preexisting DNA cleavage. Because the hydroxyl radical generally decays quite rapidly, less than 5% of the DNA is usually cleaved in the process. Thus, low preexisting background cleavage can lead to signal-to-noise problems. We have found that it is important to use high-quality DNA purified twice over a CsCl density gradient. Furthermore, when using the alkali lysis method, the final DNA pellet after isopropyl alcohol-precipitation must be neutralized with Tris base prior to CsCl density centrifugation. Once high-quality DNA is in hand, an end-labeled DNA fragment is prepared. Often it is necessary to perform a DNase I footprint under the hydroxyl-radical reaction conditions to ensure that the protein of interest binds under those conditions. In fact, titration of the time of radical cleavage at different concentrations of protein may be useful in the event the protein has inhibitors like glycerol present. Once these preliminary steps are complete, the hydroxyl-radical footprint can be performed in a single day. The products can be analyzed via autoradiogram the following day.

OUTLINE

- Step 1: Prepare buffers (30 minutes)
- Step 2: Hydroxyl-radical protection (5–6 hours)
- Step 3: Gel electrophoresis and autoradiography
- Step 4: Data analysis

PROTOCOL: HYDROXYL-RADICAL FOOTPRINTING**Procedure**

CAUTIONS: *Ammonium sulfate, Bromophenol blue, Ethanol, Formamide, H₂O₂, KCl, KOH, MgCl₂, Xylene. See Appendix I.*

Step 1: Prepare buffers*Binding buffer:*

20 mM HEPES-KOH (pH 7.5)
 50 mM KCl
 5 mM MgCl₂
 100 µg/ml BSA
 5 µg/ml poly(dI-dC)
 (use 50–100 µl per reaction)

Formamide dye mix:

98% deionized formamide
 10 mM EDTA (pH 8)
 0.025% xylene cyanol FF
 0.025% bromophenol blue

Step 2: Hydroxyl-radical protection

1. Add 10–50 femtomoles of ³²P-labeled DNA fragment and saturating concentrations of recombinant DNA-binding protein to 50–100 µl of binding buffer.
2. Incubate for 15 minutes at room temperature or optimal binding temperature determined by DNase I footprinting, for example.
3. Add 10 µl of freshly diluted 100 µM Fe(NH₄)₂SO₄ and 10 µl of freshly diluted 200 µM EDTA (do not dilute into a glass container; use an eppendorf microfuge tube).
4. Add 10 µl of 10 mM ascorbic acid and 10 µl of 0.3% H₂O₂ solution, both freshly prepared from stocks. Incubate for 2 minutes at room temperature.
5. Terminate the reaction by adding 100 µl of:
 - 5 M ammonium acetate
 - 5 mM thiourea
 - 10 mM EDTA
 - 10 µg of tRNA
 The thiourea should be added fresh from a concentrated stock.
6. (Optional) Phenol-extract.

Step 3: Gel electrophoresis and autoradiography

1. Ethanol precipitate the DNA. Wash with 80% ethanol, dry, and resuspend in formamide dye mix.
2. Denature the DNA by incubation at 95°C for 2 minutes.
3. Load directly onto a 10% or 12% polyacrylamide/urea gel and electrophorese 4–6 hours, depending on fragment size.
4. Dry gel under vacuum.
5. Expose dried gel to film or phosphorimager.

ADDITIONAL CONSIDERATIONS

1. The hydroxyl radical should generate a ladder of DNA fragments approximately equal in intensity down the gel (except in stretches where DNA bending or twisting is occurring). The ladder should be dependent on the addition of peroxide, ascorbate, and Fe-EDTA. Appropriate controls lacking these chemicals must be performed in parallel.

As described above, some proteins are not compatible with binding in peroxide. The relative balances of peroxide, ascorbate, and Fe-EDTA can easily be manipulated to accommodate alternative reaction conditions (i.e., lower peroxide but raise ascorbate). These issues are covered in detail by Dixon et al. (1991).

To determine whether your protein is affected by the radical conditions, a DNase I footprinting reaction can be performed in the presence of the chemicals necessary for generating the radical. If certain chemicals or combinations thereof inhibit binding in a footprinting assay, the conditions must be adjusted accordingly.

2. The radical can also detect DNA alterations such as bending, because these are characterized by a series of alternating 5-bp regions of reduced and enhanced cleavage due to narrowing and expanding of the minor groove through the bend. The radical also detects other unusual DNA structures as discussed by Dixon and colleagues (1991).
3. There are other variations of this technology including the missing nucleoside experiment. DNA is first treated with hydroxyl radical to remove nucleosides randomly on average of one per molecule. The ability of a protein to bind to the gapped DNA is measured by EMSA. The bound and unbound DNAs are excised, purified, and fractionated on a polyacrylamide/urea gel. The bound fraction is enriched for molecules missing bases that are not important for protein binding, while the unbound fraction is enriched in molecules missing bases important for binding. The positions of the bases are identified on a polyacrylamide/urea gel, alongside a sequencing ladder of the fragment. The technique can be used to scan a DNA molecule at single-nucleotide resolution in a single experiment (Hayes and Tullius 1989).

TROUBLESHOOTING

No cleavage

Possible cause: Dilute solutions of ascorbic acid, peroxide, and EDTA can all become inactivated over short times. Even the dry reagents, particularly ascorbic acid, can go bad in moist climates.

Solution: If no cleavage ladder is observed, try using fresh chemicals. Also, ensure that neither the activator preparation nor buffers contain more than 1% glycerol. Even 1% glycerol, however, is inhibitory.

No specific cleavage

Possible cause: DNA not prepared carefully enough.

Solution: If a preexisting ladder exists in the absence of cleavage, use greater care in preparing the plasmid DNA and the DNA fragment. Try re-preparing it. Also, prepare fresh, deionized formamide loading dye mix.

PROTOCOL 13.3

Phosphate Ethylation Interference Assay

The ethylation interference assay was originally developed by Walter Gilbert and colleagues to study RNA polymerase binding to *E. coli* promoters (Siebenlist et al. 1980; Siebenlist and Gilbert 1980).

The technique (reviewed in Manfield and Stockley 1994) involves using ethylnitrosourea to ethylate phosphates within a ^{32}P -end-labeled DNA containing a protein recognition site. This modification neutralizes the phosphate charge and places a bulky ethyl group along the edge of the minor groove (Fig. 13.1). The ethylated phosphates interfere with proteins coming into close proximity to the minor groove or phosphate backbone of the binding site. Most modifications will have little effect because only a subset of phosphates within the site will interfere and most phosphates flanking the site will not interfere (see Fig. 13.13).

The bound and unbound DNAs are then separated by EMSA (or filter binding, immunoprecipitation, etc.). The gel is autoradiographed, and the bound and free DNA populations are excised from the gel and purified from the gel slices. The unbound fraction will be enriched in DNA molecules containing modifications at positions that interfere with protein binding. In contrast, the bound fraction will contain modifications that do not interfere with binding of the protein. Ethylation makes DNA susceptible to piperidine-induced cleavage. The cleavage products are then fractionated on a polyacrylamide/urea gel alongside a sequencing ladder.

Note that cleavage with alkali can leave the phosphate attached to either the 5' or 3' breakpoints. Thus, the cleavage products often migrate as closely spaced doublets on a gel.

TIME LINE AND ORGANIZATION

The ethylation interference assay can be performed in 2 days. Most of the organization involves first radiolabeling the fragment and determining by EMSA an amount of protein that is barely saturating (see Protocols 13.5 and 13.6). Higher amounts of DNA are used in the initial EMSA to ensure conditions necessary for adequate recovery of DNA during the actual procedure. The DNA fragment is then ethylated, purified, and subjected to another round of EMSA, whereupon the bound and unbound DNAs are excised and purified. The purified DNAs are then cleaved with piperidine and fractionated on a polyacrylamide urea sequencing gel alongside a sequencing ladder.

OUTLINE

- Step 1: Prepare buffers (1 hour)
- Step 2: Ethylation interference assay (10 hours)
- Step 3: Cleave ethylated residues (1 hour)
- Step 4: Fractionate DNA from bound and unbound fractions on polyacrylamide/urea gel (3 hours)

PROTOCOL: ETHYLATION INTERFERENCE ASSAY

PROCEDURE

CAUTIONS: *DTT, Ethanol, Ethylnitrosourea, Formamide, KCl, MgCl₂, NaOH, NaPO₄, Sodium cacodylate. See Appendix I.*

Step 1: Prepare buffers

Buffer A:

10 mM Tris-HCl (pH 8.0)
10 mM MgCl₂
100 mM KCl
1 mM DTT
0.1 mM EDTA

0.05 M Sodium cacodylate (pH 8.0)

Step 2: Ethylation interference assay

1. Add 100 μ l of sodium cacodylate (pH 8.0) and 100 μ l of 95% ethanol saturated with ethyl nitrosourea to 10 picomoles of ³²P-end-labeled DNA fragment.
2. Incubate for 1 hour at 50°C.
3. Precipitate DNA twice with 10 μ l of 5 M ammonium acetate and 2.5 volumes of ethanol.
4. Resuspend the pellet in 100 μ l of Buffer A.
5. On ice, add saturating concentrations of DNA-binding protein to 50,000 cpm of ethylated DNA fragment under standard binding conditions. Incubate at 30°C for 15 minutes.
6. Carefully load and electrophorese the protein–DNA mixtures through a pre-run 4.5% native polyacrylamide gel (Sambrook et al. 1989, pp. 13.45–13.57).
7. Expose gel to film; take care that the gel and the film are oriented correctly (see Chapter 4, Protocol 4.3, Step 3 for isolating a radiolabeled DNA fragment from a polyacrylamide gel).
8. Align gel with autoradiograph and excise bands corresponding to bound and unbound fractions with a razor blade.
9. Recover the DNA fragments from the gel by electroelution (see Protocol 13.6, Step 3).
10. Ethanol-precipitate and dry the DNA pellets.

Step 3: Cleave ethylated residues

1. Resuspend the pellets from the bound and unbound fractions in 10 mM NaPO₄ (pH 7.0), 1 mM EDTA.
2. Add 2.5 μ l of 1 M NaOH. Incubate at 90°C for 30 minutes. During this step, the ethylated DNA backbone is cleaved with NaOH.

3. Add 1:1 formamide loading dye (see Protocol 13.1) to samples, and load onto a 10% polyacrylamide/urea gel alongside a sequencing ladder.
4. Dry gel under vacuum.
5. Expose the gel to film and perform autoradiography or phosphorimager analysis.

ADDITIONAL CONSIDERATIONS

1. The initial ethylation frequency must be determined empirically (see Fig. 13.2). Alter the amounts of ethylnitrosourea and/or the time of modification. Analyze the cleavage of naked DNA to ascertain conditions for less than one modified phosphate per DNA (see Box 13.4).
2. Remember that in interference techniques you need not obtain 100% binding to extract useful information. The bound lane, although enriched in ethylated phosphates, which do not interfere with binding, is depleted of ethyl phosphates, which do interfere. Simply comparing the cleaved DNAs from the bound lane with a standard ethylation ladder and determining the positions of missing bands in the bound lane can yield important information.
3. Ethylation conditions are not compatible with direct binding experiments and thus there is no ethylation protection procedure.
4. The data obtained from ethylation interference can often be useful in determining backbone contacts and can be compared and contrasted with hydroxyl radical data as discussed in Protocol 13.2.

TROUBLESHOOTING

The ethylated DNA does not gel shift when protein is added

Possible cause: This may indicate that the DNA is highly overethylated.

Solution: Try lower concentrations of ethylnitrosourea or shorter modification times.

Possible cause: The DNA may not have been purified adequately from the ethylnitrosourea.

Solution: Try re-precipitating the DNA once or twice with ethanol. Make sure that pH of resuspended DNA is near 7.

The bound and unbound DNA fractions from EMSA display the same band distribution when isolated and compared on a polyacrylamide/urea gel

Possible cause: This result is often indicative of preexisting banding or cleavage pattern, resulting from impure or old DNA.

Solution: Prepare a fresh plasmid prep and take the precautions outlined as per the hydroxyl radical footprinting procedure (see Protocol 13.2).

PROTOCOL 13.4

Methylation Interference Assay

The methylation interference and protection techniques were initially developed by Walter Gilbert and colleagues to study the interaction of *E. coli* RNA polymerase with the major groove of the DNA. The chemistry was an offshoot of that used in the chemical method of DNA sequencing invented by Maxam and Gilbert. Methylation is now widely employed as the major technique for studying interaction of proteins with major groove guanines, although it also detects adenines in the minor groove and cytosines in single-stranded DNA.

There are two versions of the method. In the protection technique, a protein is prebound to a ^{32}P -end-labeled DNA containing a recognition site and the protein–DNA complex is subjected to attack by DMS. Bound protein protects guanines (Gs) in the major and adenines (As) in the minor grooves from methylation and subsequent cleavage by piperidine. Although the protection technique works smoothly for many proteins, it has several drawbacks. Because DMS modifies certain buffers (e.g., Tris), the protection reaction was historically performed in cacodylate, which is rarely used as a standard protein-binding buffer these days. Additionally, some proteins are directly modified by DMS, and these modifications can disrupt their binding to DNA (e.g., TFIID). Furthermore, unlike interference techniques, protection techniques require nearly stoichiometric binding (~90%) to observe footprints and gain information on the binding site. For these reasons and others that are described below, the methylation interference technique has been widely employed in place of protection. It is certainly one of the highest-resolution methods for measuring the bases involved in sequence-specific recognition by proteins.

The technique involves first modifying a ^{32}P -end-labeled DNA containing a recognition site (on average once per DNA molecule; see Figure Box 13.4) and then binding the protein. Certain modifications prevent the protein from binding because they are at positions where the protein comes into close proximity with the DNA. Most modifications, however, have no effect.

The bound and unbound DNAs are then separated by EMSA (or filter binding, immunoprecipitation, etc.). The gel is autoradiographed and the bound and free DNA populations are excised from the gel and purified from the gel slices. The unbound fraction will be enriched in DNA molecules containing modifications at positions that interfere with protein binding. In contrast, the bound fraction will contain modifications that do not interfere with binding of the protein. Methylation weakens the nucleotide and makes it susceptible to piperidine-induced depurination. This leads, in turn, to scission of the DNA backbone. The cleaved fragments can then be fractionated on a sequencing gel alongside A+G and C+T chemical sequencing ladders to identify the affected positions.

TIME LINE AND ORGANIZATION

The DMS footprint can be performed in 2 days. On day 1, the DNA probe is methylated, subjected to EMSA with the protein of interest, isolated, and purified. On day 2, the methylated DNA probes are cleaved with piperidine, and the fragments are analyzed on a sequencing gel followed by autoradiography or phosphorimager analysis.

* Adapted from Ausubel et al. 1994, p. 12.3.1.

OUTLINE

- Step 1: Prepare buffers (1 hour)
- Step 2: Prepare methylated DNA probe (2 hours)
- Step 3: Bind methylated probe to protein and isolate free and bound DNA by EMSA (5 hours)
- Step 4: Cleave DNA with piperidine (3 hours)
- Step 5: Analyze cleavage ladders on sequencing gels (4 hours)

PROTOCOL: METHYLATION INTERFERENCE ASSAY

Procedure

CAUTIONS: *DMS, Ethanol, Formamide, β-Mercaptoethanol, MgCl₂, Piperidine, PMSF, Sodium cacodylate. See Appendix I.*

Step 1: Prepare Buffers

DMS reaction buffer:

50 mM sodium cacodylate (pH 8)
1 mM EDTA (pH 8) in dH₂O

Prepare 50 ml
Store at 4°C

DMS stop buffer:

1.5 M sodium acetate (pH 7)
1.0 M β-mercaptoethanol in dH₂O

Prepare 50 ml fresh

0.3 M Na acetate/ 0.1 mM EDTA:

Volume to use
5 ml of 3 M sodium acetate
10 μl of 0.5 M EDTA
pH solution to pH 5.2

dH₂O to 50 ml

Step 2: Prepare methylated DNA probe

1. Prepare a ³²P-end-labeled DNA probe (see Protocol 13.6 for a detailed protocol on how to end-label DNA).
2. Suspend approximately 10⁶ cpm of probe in 5 to 10 μl of 10 mM Tris, pH 7.6, 1 mM EDTA.
3. Add 200 μl of DMS reaction buffer.
4. Add 1 μl of neat DMS in a fume hood.
5. Mix the tubes well by vortexing and incubate for 5 minutes at room temperature.

6. Add 40 μl of DMS stop buffer to the reaction.
7. Add 1 μl of 10 mg/ml tRNA and 600 μl of 95% ethanol. Mix and incubate for 10 minutes in a dry ice/ethanol bath. Centrifuge for 10 minutes at 14,000g in a microfuge at 4°C. Carefully remove the supernatant with a drawn-out pasteur pipet and dispose in liquid DMS waste.
8. Resuspend the pellet in 250 μl of 0.3 M sodium acetate/1 mM EDTA. Keep the tube on ice and add 750 μl of 95% ethanol. Mix and re-precipitate.
9. Repeat ethanol precipitation again exactly as in the previous step. Resuspend the pellet in 250 μl of 0.3 M sodium acetate/1 mM EDTA. Keep the tube on ice and add 750 μl of 95% ethanol, mix, and ethanol-precipitate.
10. Carefully remove the supernatant with a drawn-out pasteur pipet, wash the pellet with 80% ethanol, and microcentrifuge it for 10 minutes at 14,000 rpm. Again, carefully remove the supernatant with a drawn-out pasteur pipet and let the tube air-dry completely.
11. Measure the pellet for Cerenkov counts in a scintillation counter to determine cpm.
12. Resuspend the pellet in TE buffer at 20,000 Cerenkov cpm/ μl .

Note: The DNA may be difficult to resuspend. Heat, vortex, and pipet the DNA up and down to solubilize.

Step 3: Bind methylated probe to protein and isolate probes

1. Prior to initiating this analysis, a study must be performed to optimize the amount of protein necessary to saturate an unmodified DNA fragment in an EMSA reaction. The DMS interference experiments will then utilize amounts of protein that are just barely saturating. Set up a series of three or four DNA-binding reactions in 0.5-ml siliconized eppendorf tubes (this procedure is the same one used for EMSA and DNase I footprinting). The reaction is scaled up slightly to increase the number of cpm, and multiple reactions are performed side by side to ensure uniformity in the result and also to again increase the number of cpm.

For each reaction, mix:

DNA-binding protein	1.00 μl
^{32}P -end-labeled DNA probe (50 fmole/ μl)	0.50 μl
poly dI:dC (1 $\mu\text{g}/\mu\text{l}$)	0.20 μl
BSA (50 $\mu\text{g}/\mu\text{l}$)	0.05 μl
β -mercaptoethanol (14 M)	0.05 μl
MgCl_2 (0.1 M)	0.75 μl
PMSF (0.1 M)	0.05 μl
dH_2O	3.80 μl
Buffer D ⁺ (0.1 M KCl)	6.60 μl
Total volume	13.0 μl

2. Incubate at 30°C for 1 hour.
3. Load binding reactions on a native (nondenaturing) 4.5% polyacrylamide gel as in the EMSA Protocol 13.5 (p. 493).

4. Autoradiograph the gel and cut out the bands corresponding to the protein–DNA complex (bound) and free probe. See Protocol 4.2, Step 3, on how to isolate a radioactive fragment from a polyacrylamide gel.
5. Purify the DNA from each of the gel slices by electroelution (Protocol 13.6, Step 3).

Step 4: Cleave modified DNA with piperidine

1. Resuspend the pellets from bound and unbound samples in 100 μ l of 1 M piperidine in a chemical hood.
2. Incubate the reaction mix at 90°C for 30 minutes.
Note: Lock the tubes with lid locks, so that they do not pop open!
3. After incubation, place tubes on dry ice.
4. Make holes in the tops of the tubes with a large-gauge needle and dry the samples in a vacuum evaporator for 1 hour or until dry. Add 100 μ l of distilled H₂O. Freeze and dry again. Repeat freezing and drying. Measure sample for Cerenkov counts to determine cpm.

Step 5: Analyze fragments on DNA sequencing gels

1. Based on the Cerenkov counts, add sufficient formamide sequencing gel-loading buffer to the pellet so that 1–2 μ l will contain 3000 cpm of the sample to be loaded. Heat-denature the samples for 5 minutes at 90°C. Quickly chill on ice.
2. For an overnight exposure with intensifying screens, 3000 cpm/per lane is generally sufficient. It is critical to equalize the number of counts applied for the bound and free probe to allow accurate comparison between samples.
3. Load the samples from the free probe and the bound complex onto a 6–12% (depending on size of fragment) polyacrylamide/urea sequencing gel (see Sambrook et al. 1989, pp. 13.45–13.57). Electrophorese the samples as for a sequencing gel and expose gel to autoradiographic film or a phosphorimager screen overnight. A Maxam-Gilbert or dideoxy sequencing ladder should be electrophoresed alongside to distinguish the Gs from the As and to assign a precise position to each modification.

ADDITIONAL CONSIDERATIONS

1. Remember that in interference techniques you do not need to obtain 100% binding to obtain useful information. The bound lane, although enriched in methylated adenines and guanines, which do not interfere with binding, is depleted in methylated guanines, which interfere. Simply comparing the cleaved DNAs from the bound lane with a standard methylation ladder prepared without protein (and not separated by EMSA) and determining the positions of missing bands (in the bound lane) can yield important information.
2. Heavy overmethylation of DNA prior to the EMSA can lead to methylation at multiple positions in a single DNA molecule. This can confuse the analysis because it will not be clear which one of the methylated bases interfered with protein binding. Usually overmethylation is apparent based on band intensity; the uncleaved DNA on the sequencing gel is absent or present in low quantities relative to the cleavage products.

TROUBLESHOOTING

No difference between bound and unbound DNA

Possible cause: The methylation works well based on the cleavage patterns but the DNA from the bound and free samples generates the same cleavage pattern. It is plausible that there are no Gs in the region bound by the protein. In some instances, adding too much protein can overcome the negative influence of the methylated G, and most of the DNA will appear in the bound lane.

Solution: In this case, simply titrate down the protein concentration until the amount of unbound DNA begins to increase.

No cleavage of Gs

Possible cause: This can be because the DNA was undermethylated in the first place.

Solution: In such a case, a titration of the amount of DMS or time of treatment must be performed to optimize modification.

Possible cause: The piperidine may have gone bad.

Solution: Try again with fresh new reagents.

Pre-existing cleavage ladder

Possible cause: The plasmid DNA may have already been depurinated extensively during preparation.

Solution: Prepare fresh plasmid DNA. Take care to pH DNA prior to CsCl₂ gradient.

PROTOCOL 13.5

Electrophoretic Mobility Shift Assays

SUMMARY

In an electrophoretic mobility shift assay (EMSA) or simply “gel shift,” a ^{32}P -labeled DNA fragment containing a specific DNA site is incubated with a cognate DNA-binding protein. The protein–DNA complexes are separated from free (unbound) DNA by electrophoresis through a nondenaturing polyacrylamide gel. The protein retards the mobility of the DNA fragments to which it binds. Thus, the free DNA will migrate faster than the DNA–protein complex. An image of the gel is used to reveal the positions of the free and bound ^{32}P -labeled DNA (Kerr 1995).

EMSA is one of the most sensitive methods for studying the binding properties of a protein for its site (see Chapter 8). It can be used to deduce the binding parameters and relative affinities of a protein for one or more sites or for comparing the affinities of different proteins for the same sites (Fried 1989). It can also be employed to study higher-order complexes containing multiple proteins. For example, a single protein binding to a single site would generate one predominant shifted complex. If another protein bound on top of that, it would generate an additional shift or “supershift.” Finally, EMSA can be used to study protein- or sequence-dependent DNA bending (Crothers et al. 1991). The many uses of EMSA are described throughout the text. We have designed the protocol for a pure recombinant protein but describe modifications for cruder preparations in additional considerations.

A summary of the theory and other practical aspects of EMSA methodology are found in Fried and Bromberg (1997). Ideally, the DNA concentration in the reaction should be below the K_d to obtain accurate and physiological binding measurements. Because most eukaryotic transcription factors bind with K_d values of 10^{-9} to 10^{-10} M, the DNA should be below that. For a 10- μl reaction, 1 fmole of fragment is $\sim 10^{-10}$ M DNA. For a 50-kD protein, 0.5 ng in a 10- μl reaction is equal to 10^{-9} M. Note that much DNA-binding activity can be lost during protein purification due to inactivating modifications (e.g., oxidation of key cysteines). Thus, 10^{-9} M protein added does not ensure 10^{-9} M active protein. For pure proteins, BSA or low concentrations of a nonionic detergent must be added to prevent binding of the protein to the plastic tube. Carrier DNA must be added because all proteins bind both specifically and nonspecifically to DNA and, in the absence of carrier DNA, the non-specific binding occurs on the fragment and can obscure specific binding. DTT is added to prevent protein oxidation during incubation, particularly of cysteines, and KCl and MgCl_2 are added to minimize charge repulsion of the phosphates in the backbone and to help stabilize protein and DNA structure.

TIME LINE AND ORGANIZATION

The first step is to prepare a ^{32}P - end-labeled DNA fragment or double-stranded oligonucleotide containing the site of interest. Next a 4–4.5% native polyacrylamide gel is poured and pre-electrophoresed to remove ammonium persulfate (APS) from the gel.

Occasionally various buffer components are included in the gel running buffer to mimic the reaction buffer, although this is generally not necessary (see Chapter 8). Standard binding reactions are then performed and loaded gently into the wells. After a 4- to 6-hour run, depending on the DNA fragment size, the gel can be fixed in a methanol/acetic acid mixture, attached to Whatman 3MM paper, and dried under a vacuum. Ideally, 1000 cpm of labeled fragment is employed in the reaction. Because the radioactivity partitions between the bound and unbound DNA (two or so bands), this amount of radioactivity allows the resulting autoradiograph to expose and bands to be detected in a few hours to overnight.

OUTLINE

Step 1: Prepare buffers

Step 2: Electrophoretic mobility shift assay (EMSA)

PROTOCOL: ELECTROPHORETIC MOBILITY SHIFT

Procedure

CAUTIONS: *Acetic acid (concentrated), Acrylamide, Bromophenol blue, DTT, Glycerol, KCl, KOH, β -Mercaptoethanol, Methanol, $MgCl_2$, PMSF, Radioactive substances, TEMED. See Appendix I.*

Step 1: Prepare buffers

Gel fixing solution:

1 liter of fixative:
700 ml of dH₂O
200 ml of methanol
100 ml of acetic acid

Buffer D⁺ (0.1 M KCl):

Final concentrations
20 mM HEPES-KOH (pH 7.9)
20% v/v glycerol
0.2 mM EDTA
0.1 M KCl
0.5 mM PMSF
1 mM DTT (add PMSF and DTT just before use)

Electrophoresis buffer:

0.5× TBE
1% glycerol

Step 2: Electrophoretic mobility shift assay (EMSA)

1. Pour a 40-ml 4.5% native acrylamide gel (1- to 1.5-mm spacers).

final concentrations

6.0 ml of acrylamide mix (30%, 29:1 acrylamide:bisacrylamide)	4.5% acrylamide mix
4 ml of 5× TBE	0.5× TBE
2 ml of glycerol (20% v/v)	1% glycerol
28 ml of H ₂ O	
300 μl of APS (10%)	0.075% APS
Add 30 μl of TEMED just before pouring.	

The gel must be pre-run at 10 mA for 2 hours.

2. Set up binding reactions in 0.5-ml siliconized eppendorf tubes.

Recombinant protein ^a (0.5–100 ng)	1.00 μl
³² P-DNA template ^a (ideally 1 fmole)	1.00 μl
poly (dI:dC) ^a (1 μg/ μl)	0.20 μl
BSA (50 μg/μl)	0.25 μl
DTT (0.1 M)	0.10 μl
MgCl ₂ ^b (0.1 M)	0.75 μl
Buffer D ⁺ (0.1 M KCl)	6.70 μl
<hr/>	
Total volume	10.0 μl

^a These amounts must be titrated.

^b Usually omitted when using crude extracts.

3. Incubate at 30°C (room temperature or ice for crude extracts) for 1 hour.
4. Load the samples directly (no dye added) onto the pre-run 4.5% native polyacrylamide gel. Very carefully layer the mix onto the bottom of the well and watch the sclieren from the glycerol–buffer interface.
Note: Add 5 μl of 10× DNA loading buffer with dyes to one lane that does not contain a reaction for use as a marker to determine how far the products have traveled in the gel.
5. Electrophorese at 10 mA for desired time (for a 30-bp fragment so that the bromophenol blue dye migrates about two-thirds of the way down the gel). When the gel run is complete, carefully pour out the buffer into the sink and remove the gel from the apparatus. Remove comb and split plates. Leave the gel attached to one plate.
6. *Optional:* Fix the gel in methanol/acetic acid for 15 minutes and then place on top of two sheets of Whatman 3MM paper (cat. #3030-917). Cover the top of the gel with Saran Wrap and dry on a gel dryer for 1 hour at 70°C.
7. Expose the dried gel to autoradiography film or phosphorimager overnight.

ADDITIONAL CONSIDERATIONS

1. For binding measurements on short DNA sites, perform these reactions with double-stranded oligonucleotides. When using restriction fragments, the effective limit is generally around 250 bp (see Chapter 8).
2. Large protein complexes greater than 1 Megadalton in mass are difficult to resolve and migrate quite slowly in the native polyacrylamide gels. For analyzing larger complexes, use either agarose or acrylamide-agarose composite gels.

3. Cruder preparations of protein, including nuclear extract, must be widely titrated to observe specific binding; a titration of additional carrier DNA is usually essential in such cases. When crude extracts are being used, often microgram amounts of extract or less are used. Many proteins in crude preparations bind nonspecifically to DNA and generate shifted complexes. To ensure specific binding, fragments bearing mutations in the site or competition with specific and nonspecific oligonucleotide competitors must be employed. See Chapter 8 for a discussion of the relevant parameters. Antibodies against the protein of interest can also be employed in supershift experiments to determine specificity. In such cases, however, a reaction containing the antibody itself should be performed and compared with pre-immune serum or an antibody prepared similarly.
4. Optimization of binding generally involves titrating the buffer pH, salt, and magnesium concentrations as well as the balance of carrier DNA and specific labeled fragment (see Fried 1989).
5. Some proteins simply do not bind well in gel-shift experiments and are more compatible with DNase I footprinting or another methodology (see Chapter 8). This can be due to high K_d values (weak binding) causing extensive dissociation during gel electrophoresis. Also recall that once the protein is in the gel, the buffer is TBE, which may not be compatible with binding. Simply testing binding in TBE can help determine whether this is the case. Other low-salt running buffers can be employed in place of TBE, including HEPES and TAE. Sometimes divalent ions can be added to the gel and buffers. In such cases, however, the conductivity of the gel must be kept to a minimum to prevent the gel from overheating, and sometimes if the buffering capacity of an electrophoresis buffer is low, the buffer should be recirculated among the top and bottom wells to maintain pH. Fried and Bromberg (1997) have discussed some relevant parameters that influence the stability of protein–DNA complexes during gel electrophoresis.

TROUBLESHOOTING

No binding

Possible cause: No binding generally indicates the protein is not concentrated, does not bind the site, or the buffer conditions are not compatible with protein binding during electrophoresis.

Solution: Try greater concentrations of protein, alter buffer pH and salt concentrations, try alternative electrophoresis running buffers, run gel at 4°C in cold room.

Smeary gels

Possible cause: Smeary gels may indicate protein dissociation during the electrophoresis, too much protein in the reaction, or electrophoresing the gel at too high a temperature.

Solution: Electrophorese the gel more slowly, reduce protein concentration, try running the gel in the cold room, alter the buffer conditions, or add more competitor DNA.

Too many bands

Possible cause: Too many bands are generally caused by nonspecific binding.

Solution: Try adding less protein and more carrier DNA.

PROTOCOL 13.6

Preparation of ^{32}P -end-labeled DNA Fragments

The generation of a uniquely ^{32}P -end-labeled DNA fragment is essential for performing DNA-binding experiments such as DNase I footprinting and ethylation interference. We briefly describe a protocol for end-labeling a restriction fragment (see also Chapter 4).

TIME LINE AND ORGANIZATION

For a plasmid DNA bearing a region containing the binding site of interest, the protocol involves cleaving with a single restriction endonuclease, which generates a 5' overhang containing a phosphate. This is generally necessary for both common forms of fragment end-labeling using either phosphorylation with polynucleotide kinase or "filling in the end" with DNA polymerases such as Klenow fragment. We focus on the phosphorylation reaction. The phosphate is removed with calf intestinal phosphatase, bacterial alkaline phosphatase, and later the free 5'OH generated by this manipulation is phosphorylated with polynucleotide kinase and [γ - ^{32}P]ATP. This in turn generates a plasmid now labeled at each end with γ - ^{32}P . To generate a uniquely end-labeled DNA fragment, the labeled plasmid is heat-treated to inactivate any remaining kinase and then recleaved with the second endonuclease, releasing a short DNA fragment and a longer vector fragment. The DNA fragment is purified from the labeled vector on a 5–8% native polyacrylamide gel. The molar amount of plasmid DNA must be below the amount of ATP added to the reaction and the ATP must be of high specific activity to generate a fragment labeled to the extent necessary for many DNA-binding experiments.

In contrast, an oligonucleotide primer is generally unphosphorylated at the 5' end and can be directly phosphorylated with kinase, bypassing the phosphatase step.

Preparation and labeling of DNA restriction fragments typically is performed in 1–2 days.

OUTLINE

 ^{32}P -end-labeling of a DNA fragment (2 days)

Step 1: Cleavage and labeling of the plasmid DNA (4 hours)

Step 2: Generating a uniquely labeled DNA fragment (5 hours)

Step 3: Electroelute the labeled DNA fragment (3 hours)

PROTOCOL: ³²P-END-LABELING OF A DNA FRAGMENT

PROCEDURE

CAUTIONS: *Acrylamide, Ammonium persulfate, Bromophenol blue, Chloroform, Ethanol, Phenol, Radioactive substances, TEMED, Xylene. See Appendix I.*

Step 1: Cleavage and labeling of the plasmid DNA

1. Digest 10 μg of plasmid DNA in a 1.5-ml microfuge tube with an enzyme generating a 5'-phosphate overhang. Typically, 1 μg of a 4-kb plasmid is equal to 380 fmoles of DNA. When cleaved once, it generates 760 fmoles of ends. Thus, 10 μg is equal to 7.6 pmoles of DNA ends. We label the DNA with 100 μCi or 20 pmoles of high-specific-activity ATP (5000–7000 Ci/mmol).

DNA (10 μg)	20.0 μl
10x restriction enzyme buffer 1	5.0 μl
restriction enzyme 1	5.0 μl
calf intestinal phosphatase (10 units)	0.5 μl
dH ₂ O	19.5 μl
<hr/>	
Total volume:	50.0 μl

2. Incubate at 37°C for 2 hours to cleave the plasmid and to phosphatase the DNA ends.
3. Add 1 μl of Proteinase K (stock at 10 mg/ml) and SDS to 0.1% and incubate at 55°C for 15 minutes to digest residual phosphatase and restriction enzyme.
4. Add 50 μl of TE (10 mM Tris pH, 7.9, 1 mM EDTA) and 100 μl of phenol/chloroform (1:1) and vortex gently. Centrifuge for 1 minute in a microfuge at 14,000g, then carefully remove 90 μl of the aqueous (top) phase and transfer to a new 1.5-ml eppendorf tube. Do not transfer any of the interphase; instead remove less of the aqueous phase. This manipulation effectively removes all of the phosphatase.
5. Add 10 μl of 3 M sodium acetate and 250 μl of 95% ethanol to the aqueous phase. Mix well and place on dry ice for 15 minutes. Collect the DNA precipitate as a pellet by centrifugation in a microfuge at 14,000g for 15 minutes. Carefully remove the ethanol using a drawn-out glass pasteur pipet.
6. Wash the pellet by adding 100 μl of 80% ethanol, centrifuge in microfuge for 1 minute, remove the ethanol with drawn-out pipet and air-dry completely.
7. Resuspend the dried DNA pellet in 16 μl of dH₂O by gently pipetting up and down several times. ³²P-end-label with polynucleotide kinase and [γ-³²P]ATP.

Mix:	
DNA	16.0 μl
10x One-Phor-All buffer (Amersham Pharmacia Biotech Inc. cat. # 27-0901-02)	3.0 μl
[γ- ³² P]ATP 5-7000 Ci/mmol (10 mCi/ml aqueous)	10.0 μl
polynucleotide kinase (10 U/ μl)	1.0 μl
<hr/>	
Total volume	30.0 μl

8. Mix very gently and then incubate at 37°C for 15 minutes. Exercise maximum precautions and shielding to prevent exposure to or spilling of ³²P. This reaction end-labels the free 5'OH ends generated by restriction endonuclease cleavage and phosphatase treatment (Steps 1–5).
9. After the end-labeling reaction, heat-inactivate the kinase by incubating at 65°C for 15 minutes.
10. Phenol/chloroform-extract, ethanol-precipitate, and wash the DNA pellet as in Steps 4–6. Keep a 50-ml conical tube nearby to store the radioactive drawn-out pipet.

Step 2: Create a uniquely labeled DNA

1. Resuspend the pellet in 10 µl of ddH₂O and digest with the second enzyme.

Mix:

DNA	10.0 µl
10× restriction enzyme 2 buffer	5.0 µl
restriction enzyme 2 (10 units)	5.0 µl
dH ₂ O	30.0 µl
<hr/>	
Total volume	50.0 µl

2. Digest DNA at 37°C for 3 hours.
3. Add 5 µl of 10× DNA loading buffer and use sequencing pipet tips to load the mixture onto a native 8% polyacrylamide gel (for fragments in the 100-bp range) and electrophorese in 1× TBE buffer. On an 8% gel, the bromophenol blue dye migrates with the 45-bp marker, and the xylene cyanol dye runs with the 160-bp marker.
4. Prepare an 8% native acrylamide gel. Mix the following ingredients in a 50-ml conical disposable tube.

Mix:

40% acrylamide (29:1)	6 ml
5× TBE buffer	6 ml
dH ₂ O	18 ml
<hr/>	
Total volume	30 ml

Add 300 µl of 10% APS

Add 30 µl of TEMED

Mix well and rapidly pour into gel plates. Insert comb and clamp the top of gel to hold the comb tightly. Allow the gel to polymerize for 1 hour, carefully remove the comb, and prerun the gel for 1 hour at 200 V to remove APS.

5. Load the sample slowly into gel wells by layering, and electrophorese until the bromophenol blue has migrated approximately halfway down the gel.

Step 3: Electroelute the labeled DNA fragment

1. Very carefully remove the gel from the vertical rig and pour radioactive buffer from lower chamber into radioactive waste container. Rinse the remaining gel apparatus with water in a sink. Rinse the plates with water and then carefully separate them using a spatula. Do this on bench-top paper behind a shield. Leave the gel on one plate and carefully cover the plate with plastic Saran Wrap. Bring the gel to the dark room and

expose it to XAR-5 photographic film (Kodak) for 3 minutes to locate position of radiolabeled fragment. Use fluorescent dye to mark plastic wrap around gel (dot in three places) or flash the gel three times with a camera flash apparatus. The latter generates an outline of the gel and plate on the film, and the former generates a series of black dots. Either method will be necessary to align the autoradiograph later against the gel to excise the band. Develop the film.

2. Align the film and gel using fluorescent markers or the flash outline. Excise the gel slice that corresponds to the probe using a razor blade and tweezers. Cut through the Saran Wrap.
3. Place the gel slice in dialysis tubing (10-mm flat diameter, 12,000–14,000 Dalton MWCO approx. 2 inches long) with 500 μ l of 1 \times TBE and electroelute the DNA from the gel for 1 hour at 100 V. Place the dialysis tubing containing the gel slice into a horizontal agarose gel chamber containing 1 \times TBE. After electroeluting for 1 hour at 100 V, switch the direction of the electrodes and electroelute for 2 minutes. (This will elute any DNA off the tubing and back into the buffer.)
4. Use a 1-ml pipetman to remove buffer containing the eluted DNA fragment from the dialysis tubing and place in a 1.5-ml microfuge tube.
5. Extract the solution with an equal volume of phenol/chloroform.
6. Add 1/10th volume of 3 M sodium acetate and 2 volumes of 95% ethanol and precipitate DNA as above.
7. Resuspend the final pellet in TE at 10 K cpm/ μ l (as judged by a Geiger counter).
8. Use a scintillation counter to measure Cerenkov radioactivity (^3H channel) of 1 μ l of probe dotted onto a 1-cm² piece of filter paper. Compare to radioactivity of 0.01 μ Ci of the starting ATP to determine incorporation and specific activity. Typically, 10,000 cpm is employed for a DNase I footprint, but much less is required for EMSA.

ADDITIONAL CONSIDERATIONS

1. Sometimes it is necessary to cleave restriction fragments with enzymes that do not generate unique sites. This is feasible as long as the ATP is in twofold or greater excess over the actual number of DNA ends. The ideal fragment for EMSA and footprinting reactions is 50–200 nucleotides in length.
2. An alternative method used by many researchers involves using PCR to generate the end-labeled fragment. In such a case, approximately 10 ng of plasmid DNA is incubated with 10 pmole each of two primers flanking the region of interest, one of which is end-labeled. After 20–30 cycles of PCR, an end-labeled fragment of high specific activity can be generated and recovered by purification on a native acrylamide gel as described in Steps 2 and 3 (above).

TROUBLESHOOTING

No or weak labeling of DNA fragments

Possible cause: DNA may not have been added.

Solution: Check concentration of DNA stocks by A_{260} or by simply running a diagnostic agarose gel for plasmids to ensure that DNA was indeed added. If DNA was added, check kinase and buffers.

Possible cause: Old DTT can inactivate the kinase.

Solution: Fresh DTT is needed to maintain polynucleotide kinase activity.

Possible cause: Background labeling of contaminating nucleic acids. In plasmid preps contaminated with RNA the kinase will also phosphorylate RNA ends.

Solution: Removing RNA by CsCl EtBr gradients or simply rerunning a QIAGEN column can help to remove excess RNA. Do not add RNase; it creates a greater problem by increasing the number of ends.

Possible cause: Inactive kinase.

Solution: Use freshly purchased kinase.

REFERENCES

- Ausubel F.M., Brent R.E., Kingston E., Moore D.D., Seidman J.G., Smith J.A., and Struhl K. 1994. *Current protocols in molecular biology*. John Wiley and Sons, New York.
- Balasubramaniam B., Pogozelski W.K., and Tullius T.D. 1998. DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. *Proc. Natl. Acad. Sci.* **95**: 9738–9743.
- Blackwell T.K. 1995. Selection of protein binding sites from random nucleic acid sequences. *Methods Enzymol.* **254**: 604–618.
- Branden C. and Tooze J. 1991. *Introduction to protein structure*. Garland Publishing, New York.
- Brenowitz M., Senear D.F., Shea M.A., and Ackers G.K. 1986. Quantitative DNase footprint titration: A method for studying protein-DNA interactions. *Methods Enzymol.* **130**: 132–181.
- Burley S.K. and Roeder R.G. 1996. Biochemistry and structural biology of Transcription Factor IID (TFIID). *Annu. Rev. Biochem.* **65**: 769–799.
- Carey M. 1998. The enhanceosome and transcriptional synergy. *Cell* **92**: 5–8.
- Carey M., Kakidani H., Leatherwood J., Mostashari F., and Ptashne M. 1989. An amino-terminal fragment of GAL4 binds DNA as a dimer. *J. Mol. Biol.* **209**: 423–432.
- Cho Y., Gorina S., Jeffrey P.D., and Pavletich N.P. 1994. Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. *Science* **265**: 346–355.
- Crothers D.M., Gartenberg M.R., and Shrader T.E. 1991. DNA bending in protein-DNA complexes. *Methods Enzymol.* **208**: 118–146.
- Crothers D.M., Drak J., Kahn J.D., and Levene S.D. 1992. DNA bending, flexibility, and helical repeat by cyclization kinetics. *Methods Enzymol.* **212**: 3–29.
- Dignam J.D., Lebovitz R.M., and Roeder R.G. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**: 1475–1489.
- Dill K.A. 1997. Additivity principles in biochemistry. *J. Biol. Chem.* **272**: 701–704.
- Dixon W.J., Hayes J.J., Levin J.R., Weidner M.F., Dombroski B.A., and Tullius T.D. 1991. Hydroxyl radical footprinting. *Methods Enzymol.* **208**: 380–413.
- Echols H. 1986. Multiple DNA-protein interactions governing high-precision DNA transactions. *Science* **233**: 1050–1056.
- Ellenberger T.E., Brandl C.J., Struhl K., and Harrison S.C. 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: Crystal structure of the protein-DNA complex. *Cell* **71**: 1223–1237.
- Ellington A.D. and Szostak J.W. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**: 818–822.
- Fairall L., Rhodes D., and Klug A. 1986. Mapping of the sites of protection on a 5 S RNA gene by the *Xenopus* transcription factor IIIA. A model for the interaction [published erratum appears in *J. Mol. Biol.* 1987 Apr 5;194(3):581]. *J. Mol. Biol.* **192**: 577–591.
- Feng J.-A., Johnson R.C., and Dickerson R.E. 1994. Hin recombinase bound to DNA: The origin of specificity in major and minor groove interactions. *Science* **263**: 348–355.
- Fried M.G. 1989. Measurement of protein-DNA interaction parameters by electrophoresis mobility

- shift assay. *Electrophoresis* **10**: 366–376.
- Fried M.G. and Bromberg J.L. 1997. Factors that affect the stability of protein-DNA complexes during gel electrophoresis. *Electrophoresis* **18**: 6–11.
- Galas D.J. and Schmitz A. 1978. DNase footprinting: A simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5**: 3157–3170.
- Giese K., Kingsley C., Kirshner J.R., and Grosschedl R. 1995. Assembly and function of a TCR α enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev.* **9**: 995–1008.
- Gralla J.D. 1985. Rapid “footprinting” on supercoiled DNA. *Proc. Natl. Acad. Sci.* **2**: 3078–3081.
- Grange T., Bertrand E., Espinas M.L., Fromont-Racine M., Rigaud G., Roux J., and Pictet R. 1997. In vivo footprinting of the interaction of proteins with DNA and RNA. *Methods* **11**: 151–163.
- Griffith J., Hochschild A., and Ptashne M. 1986. DNA loops induced by cooperative binding of λ repressor. *Nature* **322**: 750–752.
- Grosschedl R. 1995. Higher-order nucleoprotein complexes in transcription: Analogies with site-specific recombination. *Curr. Opin. Cell Biol.* **7**: 362–370.
- Grosschedl R., Giese K., and Pagel J. 1994. HMG domain proteins: Architectural elements in the assembly of nucleoprotein structures. *Trends Genet.* **10**: 94–100.
- Ha J.H., Spolar R.S., and Record M.T. Jr. 1989. Role of the hydrophobic effect in stability of site-specific protein-DNA complexes. *J. Mol. Biol.* **209**: 801–816.
- Hayes J.J. and Tullius T.D. 1989. The missing nucleoside experiment: A new technique to study recognition of DNA by protein. *Biochemistry* **28**: 9521–9527.
- Herschman H.R. 1991. Primary response genes induced by growth factors and tumor promoters. *Annual Review of Biochemistry* **60**: 281–319.
- Hochschild A. 1991. Detecting cooperative protein-DNA interactions and DNA loop formation by footprinting. *Methods Enzymol.* **208**: 343–361.
- Hochschild A. and Ptashne M. 1986. Cooperative binding of λ repressors to sites separated by integral turns of the DNA helix. *Cell* **44**: 681–687.
- Hoover T.R., Santero E., Porter S., and Kustu S. 1990. The integration host factor stimulates interaction of RNA polymerase with NIFA, the transcriptional activator for nitrogen fixation operons. *Cell* **63**: 11–22.
- Johnson A.D., Potete A.R., Lauer G., Sauer R.T., Ackers G.K., and Ptashne M. 1981. λ Repressor and cro—Components of an efficient molecular switch. *Nature* **294**: 217–223.
- Johnson R.C. 1995. Site specific recombinases and their interactions with DNA. In *DNA-protein: structural interactions* (ed. Lilley D.M.J.), vol. 7, p. 141. Oxford University Press, New York and IRL Press at Oxford University Press, United Kingdom.
- Jordan S.R. and Pabo C.O. 1988. Structure of the lambda complex at 2.5 Å resolution: Details of the repressor-operator interactions. *Science* **242**: 893–899.
- Kaliosis B. and O’Farrell P.H. 1993. A universal target sequence is bound in vitro by diverse homeodomains. *Mech. Dev.* **43**: 57–70.
- Kerr L.D., 1995. Electrophoretic mobility shift assay. *Methods Enzymol.* **254**: 619–632.
- Kielkopf C.L., White S., Szewczyk J.W., Turner J.M., Baird E.E., Dervan P.B., and Rees D.C. 1998. A structural basis for recognition of A-T and T-A base pairs in the minor groove of B-DNA. *Science* **282**: 111–115.
- Kim S. and Landy A. 1992. Lambda Int protein bridges between higher order complexes at two distant chromosomal loci, *attL* and *attR*. *Science* **256**: 198–203.
- Kim S., Moitoso de Vargas L., Nunes-Duby S., and Landy A. 1990. Mapping of a higher order protein-DNA complex: Two kinds of long-range interactions in Lambda, *attL*. *Cell* **63**: 773–781.
- Kim T.K. and Maniatis T. 1997. The mechanism of transcriptional synergy of an in vitro assembled interferon- β enhancer. *Mol. Cell* **1**: 119–130.
- Kim J.L., Nikolov D.B., and Burley S.K. 1993a. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**: 520–527.
- Kim Y., Geiger J.H., Hahn S., and Sigler P.B. 1993b. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**: 512–520.

- Kissinger C.R., Liu B., Martin-Bianco E., Kornberg T.B., and Pabo C.O. 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8Å resolution: A framework for understanding homeodomain-DNA interactions. *Cell* **63**: 579–590.
- Klemm J.D., Rould M.A., Aurora R., Herr W., and Pabo C.O. 1994. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* **77**: 21–32.
- Koblan K.S., Bain D.L., Beckett D., Shea M.A., and Ackers G.K. 1992. Analysis of site-specific interaction parameters in protein-DNA complexes. *Methods Enzymol.* **210**: 405–425.
- Lee D.K., Horikoshi M., and Roeder R.G. 1991. Interaction of TFIID in the minor groove of the TATA element. *Cell* **67**: 1241–1250.
- Lesser D.R., Kurpiewski M.R., and Jen-Jacobson L. 1990. The energetic basis of specificity in the Eco RI endonuclease–DNA interaction. *Science* **250**: 776–786.
- Liang S.D., Marmorstein R., Harrison S.C., and Ptashne M. 1996. DNA sequence preferences of GAL4 and PPR1: How a subset of Zn2 Cys6 binuclear cluster proteins recognizes DNA. *Mol. Cell. Biol.* **16**: 3773–3780.
- Love J.J., Li X., Case D.A., Giese K., Grosschedl R., and Wright P.E. 1995. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**: 791–795.
- Luisi B. 1995. DNA-protein interaction at high resolution. In *DNA-Protein: Structural interactions* (ed. Lilley D.M.J.). Oxford University Press, New York and IRL Press at Oxford University Press, United Kingdom.
- Manfield I. and Stockley P.G. 1994. Ethylation interference. *Methods Mol. Biol.* **30**: 125–139.
- Mangelsdorf D.J. and Evans R.M. 1995. The RXR heterodimers and orphan receptors. *Cell* **83**: 841–850.
- Marmorstein R., Carey M., Ptashne M., and Harrison S.C. 1992. DNA recognition by GAL4: Structure of a protein-DNA complex. *Nature* **356**: 408–414.
- Maxam A.M. and Gilbert W. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci.* **74**: 560–564.
- Narayana N., Ginell S.L., Russu I.M., and Berman H.M. 1991. Crystal and molecular structure of a DNA fragment: d(CGTGAATTCACG). *Biochemistry* **30**: 4449–4455.
- Oliphant A.R., Brandl C.J., and Struhl K. 1989. Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: Analysis of yeast GCN4 protein. *Mol. Cell. Biol.* **9**: 2944–2949.
- Otwinowski Z., Schevitz R.W., Zhang R.G., Lawson C.L., Joachimiak A., Marmorstein R.Q., Luisi B.F., and Sigler P.B. 1988. Crystal structure of trp repressor/operator complex at atomic resolution [published erratum appears in *Nature* 1988 Oct 27;335(6193):837]. *Nature* **335**: 321–329.
- Pabo C.O. and Sauer R.T. 1992. Transcription factors: Structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**: 1053–1095.
- Pan C.Q., Finkel S.E., Cramton S.E., Feng J.A., Sigman D.S., and Johnson R.C. 1996. Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J. Mol. Biol.* **264**: 675–695.
- Pavletich N.P. and Pabo C.O. 1991. Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1Å. *Science* **252**: 809–817.
- Pomerantz J.L., Sharp P.A., and Pabo C.O. 1995. Structure-based design of transcription factors. *Science* **267**: 93–96.
- Ptashne M. 1992. *A genetic switch: Phage λ and higher organisms*, 2nd edition. Cell Press and Blackwell Scientific Publications, Cambridge, Massachusetts.
- Ptashne M. and Gann A. 1998. Imposing specificity by localization: Mechanism and evolvability [published erratum appears in *Curr. Biol.* 1998 Dec. 3;8(24):R897] *Curr. Biol.* **8**: R812–822.
- Raumann B.E., Rould M.A., Pabo C.O., and Sauer R.T. 1994. DNA recognition by β-sheets in the Arc repressor-operator crystal structure. *Nature* **367**: 754–757.
- Rebar E.J. and Pabo C.O. 1994. Zinc finger phage: Affinity selection of fingers with new DNA-binding specificities. *Science* **263**: 671–673.
- Record M.T. Jr., Ha J.H., and Fisher M.A. 1991. Analysis of equilibrium and kinetic measurements

- to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. *Methods Enzymol.* **208**: 291–343.
- Rippe K., von Hippel P.H., and Langowski J. 1995. Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem. Sciences* **20**: 500–506.
- Sambrook J., Fritsch E.F., and Maniatis T. 1989. *Molecular cloning: A laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Schwabe J.W. and Rhodes D. 1997. Linkers made to measure. *Nat. Struct. Biol.* **4**: 680–683.
- Seeman N.C., Rosenberg J.M., and Rich A. 1976. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci.* **73**: 804–808.
- Shakked Z., Guzikevich-Guerstein G., Frolov F., Rabinovich D., Joachimiak A., and Sigler P.B. 1994. Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature* **368**: 469–473.
- Siebenlist U. and Gilbert W. 1980. Contacts between *Escherichia coli* RNA polymerase and an early promoter of phage T7. *Proc. Natl. Acad. Sci.* **77**: 122–126.
- Siebenlist U., Simpson R.B., and Gilbert W. 1980. *E. coli* RNA polymerase interacts homologously with two different promoters. *Cell* **20**: 269–281.
- Sigman D.S., Kuwabara M.D., Chen C.H., and Bruice T.W. 1991. Nuclease activity of 1,10-phenanthroline-copper in study of protein-DNA interactions. *Methods Enzymol.* **208**: 414–433.
- Somers W.S. and Phillips S.E. 1992. Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by β -strands. *Nature* **359**: 387–393.
- Starr D.B. and Hawley D.K. 1991. TFIID binds in the minor groove of the TATA box. *Cell* **67**: 1231–1240.
- Suck D. 1994. DNA recognition by DNase I. *J. Mol. Recognit.* **7**: 65–70.
- Suzuki M. 1995. DNA recognition by a β -sheet. *Protein Eng.* **8**: 1–4.
- Swaminathan K., Flynn P., Reece R.J., and Marmorstein R. 1997. Crystal structure of a PUT3-DNA complex reveals a novel mechanism for DNA recognition by a protein containing a Zn₂Cys₆ binuclear cluster. *Nat. Struct. Biol.* **4**: 751–759.
- Thanos D. and Maniatis T. 1995. Virus induction of human IFN β gene expression requires the assembly of an enhanceosome. *Cell* **83**: 1091–1100.
- Todd R.B. and Andrianopoulos A. 1997. Evolution of a fungal regulatory gene family: The Zn(II)₂Cys₆ binuclear cluster DNA binding motif. *Fungal Genet. Biol.* **21**: 388–405.
- Tuerk C., MacDougal S., and Gold L. 1992. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci.* **89**: 6988–6992.
- Tullius T.D. and Dombroski B.A. 1986. Hydroxyl radical “footprinting”: High-resolution information about DNA-protein contacts and application to λ repressor and Cro protein. *Proc. Natl. Acad. Sci.* **83**: 5469–5473.
- Wang J.C. and Gjaever G.N. 1988. Action at a distance along a DNA. *Science* **240**: 300–304.
- Wang M.M. and Reed R.R. 1993. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature* **364**: 121–126.
- Watson J.D., Weiner A.M., and Hopkins N.H. 1987. *Molecular biology of the gene*, 4th edition. Benjamin/Cummings Publishing, Menlo Park, California.
- Werner M.H. and Burley S.K. 1997. Architectural transcription factors: Proteins that remodel DNA. *Cell* **88**: 733–736.
- Wissmann A. and Hillen W. 1991. DNA contacts probed by modification protection and interference studies. *Methods Enzymol.* **208**: 365–379.
- Wolberger C., Vershon A.K., Liu B., Johnson A.D., and Pabo C.O. 1991. Crystal structure of a MAT α -2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* **67**: 517–528.
- Yang J. and Carey J. 1995. Footprint phenotypes: Structural models of DNA-binding proteins from chemical modification analysis of DNA. *Methods Enzymol.* **259**: 452–468.
- Zwieb C. and Adhya S. 1994. Improved plasmid vectors for the analysis of protein-induced DNA bending. *Methods Mol. Biol.* **30**: 281–294.