

## DNA-PROTEIN INTERACTIONS

(extracted from: Current Protocols in Molecular Biology = Red Book)

For several decades, DNA-binding proteins have been studied because of their involvement in cellular processes such as replication, recombination, viral integration, and transcription. In recent years, the number of workers interested in the study of these proteins has greatly increased as the advent of recombinant DNA technology has led to the isolation of numerous biologically important genes. Many investigators are interested in how transcription of these genes is controlled in response to environmental or developmental signals, and have started to characterize the DNA sequences responsible for this regulation. This analysis has naturally led to the detection, isolation, and characterization of the proteins that bind to these regulatory sequence elements. This chapter summarizes the techniques currently used to characterize DNA-protein interactions and to isolate DNA-binding proteins.

The preparation of cell extracts from either the nucleus or cytoplasm is often the first step in studying these proteins and their interactions (UNIT 12.1). The resulting preparations can be used directly in a variety of functional studies, or as the starting point for the purification of proteins involved in gene regulation. In the past three years many advances have been made in the technologies used for detecting DNA-binding proteins and for purifying those proteins. Perhaps the most widely applied technology has been the mobility shift assay (UNIT 12.2), whereby proteins are detected by their ability to retard mobility of a labeled DNA fragment through a nondenaturing gel. This technique--originally developed to characterize the interaction of purified prokaryotic proteins with DNA--has been refined to allow detection of numerous DNA-binding proteins in crude extracts from a wide variety of cells. Because the mobility shift assay is simple and rapid, it is typically the method of choice when purifying DNA-binding proteins.

After detection of a DNA-binding protein, it is often necessary to determine its binding specificity. Guesses about the binding site can be verified by using mutated DNA fragments as either cold competitor or as labeled probe in the mobility shift assay. Alternatively, one can chemically modify the DNA template, and ask how these alterations affect binding of a specific protein. The most widely used of these modification, or "interference," techniques is methylation interference (UNIT 12.3). This technique allows direct detection of nucleotides that are in close contact with the binding protein. Here, guanine and adenine residues are methylated at an average of one modification per DNA molecule. Modified DNA molecules bound to a specific protein are then separated from those molecules that are not bound via the mobility shift assay. Molecules that are methylated at nucleotides which are important for binding will be underrepresented in bound DNA.

While methylation interference tends to be more informative than most protection techniques, DNase I protection mapping (or "footprinting") can provide another level of information (UNIT 12.4). In these procedures, protein is first bound to DNA and then the DNA is cleaved either by DNase or chemical agents. Footprinting typically reveals, without perturbing, the general region(s) of DNA to which a protein binds. Footprinting is rapid and sensitive once it is optimized for a particular interaction and can be used as a routine assay in purification. The footprinting assay can also be used as a quantitative technique to determine both the binding curves for individual proteins, as well as cooperative interactions among proteins bound to multiple sites along DNA.

After detection and preliminary characterization of a DNA-binding protein, investigators frequently wish to purify the protein. This can be accomplished using standard chromatography techniques (Chapter 10) and assaying the fractions for presence of the protein by mobility shift gel electrophoresis or footprinting. Another procedure that has proved to be very powerful in these purification schemes is the use of an affinity column containing large amounts of DNA specifying the binding site for the factor (UNIT 12.10; Kadonaga, J.T. and Tjian, R., 1986, Affinity purification of sequence-specific DNA binding proteins, Proc. Natl. Acad. Sci. U.S.A. 83: 5889-5893).

These affinity columns take advantage of the extraordinary specificity that a protein has for its cognate binding site. When such a column is used, protein is applied to it in the presence of high levels of competing, nonspecific DNA, and proteins recognizing specific sequence motifs on the affinity column partition onto those sequences as the column is loaded. In an alternative procedure, protein is first fractionated on a standard, nonspecific DNA column (e.g., DNA-cellulose), and then applied in high salt to the specific affinity column. Specific protein-DNA complexes are frequently stable to moderate salt concentrations, while many nonspecific protein-DNA complexes are disrupted in high salt. A high degree of purification can therefore be achieved using a standard DNA-cellulose column and a DNA affinity column in tandem.

Similar degrees of purification can be accomplished through the use of biotinylated DNA fragments that contain the binding site for a protein and column matrices that are coated with streptavidin (UNIT 12.6). In this procedure, a biotinylated DNA fragment containing a specific sequence motif is mixed with crude or partially purified protein and an excess of competitor DNA. As with the affinity column described above, the specific protein will partition onto the DNA fragment containing the binding site. The specific DNA fragment--as well as the attached protein--is then fished out of this mixture using streptavidin and a column matrix, and the specific protein is eluted with high salt. This protocol is relatively easy to optimize, as the success or failure of several of the steps can be monitored using mobility shift gels. It is also extremely flexible, as one single type of column matrix is compatible with any biotinylated DNA probe, and thus can be used to purify numerous different DNA-binding proteins.

To help ensure that the appropriate protein has been purified, it is useful to know the size of the protein that interacts with a specific DNA sequence. In addition, regulatory proteins can be modified in the cell in response to environmental or developmental signals. Therefore, knowing the apparent size of a DNA-binding protein under various conditions is also important as this information can provide insight into regulatory events concerning that protein.

The size of a DNA-binding protein can be determined by covalently crosslinking the protein to its regulatory sequence using UV light and resolving the protein complexes on an SDS polyacrylamide gel (UNIT 12.5). This procedure works even for impure proteins in crude extracts. Crosslinking can be accomplished by irradiating a protein solution containing a specific labeled DNA probe. Alternatively, protein-DNA complexes can first be separated on a mobility shift gel and then irradiated. This latter protocol increases confidence that the identified protein is actually part of an appropriate complex. In both protocols, it is critical to verify that the crosslinked protein interacts specifically with a given sequence motif, for example by performing competition studies with unlabeled DNA fragments.

In many instances, an important goal is to clone the gene that encodes a DNA-binding protein, thus allowing detailed genetic characterization of the protein. This can be accomplished by purifying the binding protein to homogeneity, sequencing it, and using that sequence to identify a cDNA clone (see

Chapter 6). However, despite the recent advances in purification techniques, the above approach can be extremely time consuming, particularly if the DNA-binding protein is present at very low levels. An alternative approach is to use the sequence known to be recognized by the protein to directly identify a cDNA clone (UNIT 12.7). In this protocol, a library that expresses inserted cDNAs in *E. coli* is plated out, and proteins expressed in plaques produced by recombinant phages are transferred to nitrocellulose filters. These filters are then probed with a specific labeled sequence in order to detect clones that express a given DNA-binding protein. Success of this protocol relies on the assumption that the protein as expressed in *E. coli* will be capable of specifically binding DNA. For example, if the protein binds DNA as a heterodimer, or requires a particular covalent modification to bind to its site, it will not be detectable using this approach.

Using the above mentioned techniques, one can identify and purify DNA-binding regulatory proteins. Once these proteins are pure it is important to characterize their ability to interact with DNA. A simple assay for this characterization is filter binding (UNIT 12.8). Double-stranded DNA flows through nitrocellulose, while protein or protein-DNA complexes are retained. Filter binding provides a simple and sensitive means of characterizing the interaction of purified protein with a known regulatory sequence. Competition experiments using filter binding allow rigorous determination of the relative affinities that several sequences have for one protein. In addition, kinetic measurements can be performed that allow a good estimate of the binding constant and off-rate for any given protein-DNA interaction.

Filter binding is also a useful procedure for determining the DNA-binding properties of pure protein with unknown functions. Frequently (e.g., with developmentally important genes or with oncogenes) a gene product that is biologically important is known to reside in the nucleus, but its precise function is obscure. Filter binding can be used to identify DNA sequences that interact with these proteins. If such sequences are identified by filter binding, more detailed analyses--such as footprinting or methylation interference--can be used to further characterize the binding site.

Once a cloned gene encoding a DNA-binding protein is identified, it is possible to synthesize radiolabeled protein by *in vitro* transcription and translation (UNIT 10.7). The resulting labeled protein can then be used to detect and analyze DNA-protein interaction (UNIT 12.9). The striking advantage of this approach is the possibility for constructing mutations in the cloned gene which can affect the DNA binding properties of the expressed gene.

The protocols described in this chapter can therefore be used to detect and characterize specific DNA-protein interactions, to purify specific DNA-binding proteins, and to clone the genes for these proteins. Using these protocols, one can start with a defined regulatory sequence motif and isolate the gene that encodes the factor responsible for regulation. The characteristics of the regulatory protein can be determined, and the clone of the regulatory gene can be dissected to define functional domains of the regulatory protein. These procedures have been widely applied in mammalian systems where there is very little that can be done to dissect gene regulation by classical genetic means, and have developed to the point where it is feasible, though certainly still difficult, to dissect complex regulatory loops at the molecular level.

## **Preparation of Nuclear and Cytoplasmic Extracts from Mammalian Cells**