variation on the basic protocol may be employed. In this protocol, the same biotinylated DNA fragment is used but is removed from solution directly by streptavidin-agarose (see Fig. 12.6.2).

COMMENTARY

The biotin/streptavidin purification method works well because the interaction between biotin and avidin (or avidin-like proteins) is one of the strongest known noncovalent interactions. The dissociation constant for the streptavidin-biotin complex is ~10-15 M. Avidin, from egg white, and streptavidin, from Streptomyces avidinii, are tetrameric proteins containing four high-affinity binding sites for the vitamin biotin. Since streptavidin is multivalent, it is able to serve as a bridge between the biotinylated DNA fragment and the biotin-containing resin. The strong interaction is extremely useful for purification of DNA-binding proteins, because DNA-affinity columns with streptavidin/biotin bridges can be washed under a wide variety of conditions (i.e., 2 M KCl and 1% SDS) without removing either the streptavidin or the biotinylated DNA fragment from the matrix.

Two properties of streptavidin make it more suitable than avidin for use in DNA-affinity purification. The first is that streptavidin, unlike avidin, is not a glycoprotein, and the second is that streptavidin is slightly acidic whereas avidin is basic. Therefore, streptavidin is less likely to bind nonspecifically to cellular glycoproteins and to acidically charged cell components such as nucleic acids.

Several factors make this method simple, rapid, and effective. The same binding conditions and DNA fragment used in the mobility-shift DNA-binding assay to identify a protein (UNIT 12.2) can be used to effect its purification. In addition, the binding of the protein to its DNA recognition site in solution is more efficient than protein-DNA interactions that take place on a column matrix (discussed below). Most importantly, binding in solution allows each reaction parameter to be optimized on an analytical scale by using the gel binding assay.

As an analytical technique, biotin/streptavidin DNA-affinity purification permits the direct identification of a wide variety of sequence-specific DNA-binding proteins. It has already been successfully used to identify hormone receptors, components in mRNA splicing complexes, and RNA polymerase II and III transcription factors (Haeuptle et al., 1983; Grabowski and Sharp, 1986; Chodosh et al., 1986; Kasher et al., 1986).

Another method is commonly used for purifying DNA-binding proteins, whereby DNA-protein binding interactions occur on a column matrix. In this technique, catenated DNA-binding sites are covalently coupled with cyanogen bromide to Sepharose CL-2B (Kadonaga and Tjian, 1986; Rosenfeld and Kelly, 1986). A partially purified protein fraction is combined with competitor DNA and passed through a DNA-Sepharose resin, binding the protein to the surface of the matrix. Generally, multiple passes through the column are used to effect purification, with 25- to 50-fold purification achieved with each column passage. The biotin/streptavidin technique, on the other hand, typically yields 50- to 250-fold purification.

Detection, Purification, and Characterization of cDNA Clones Encoding DNA-Binding Proteins

In this unit, an appropriate recombinant clone is detected in an expression library with a DNA recognition-site probe, purified, and shown to encode a DNA-binding domain of defined sequence specificity. The strategy described below obviates purification of a sequence-specific DNA-binding protein for the purpose of isolating its gene; it simply requires a suitable cDNA library constructed in the expression vector lgt11 and a DNA recognition-site probe. The basic protocol enables the detection and purification of clones encoding sequence-specific DNA-binding proteins. The alternate protocol describes a denaturation/renaturation procedure that can increase detection of certain clones. The support protocol provides a rapid means of characterizing the DNA-binding activities of the proteins encoded by the cloned cDNAs.

BASIC PROTOCOL: SCREENING A lgt11 EXPRESSION LIBRARY WITH RECOGNITION-SITE DNA

A clone encoding a sequence-specific protein is detected in a lgt11 library because its recombinant protein binds specifically to a radiolabeled recognition-site DNA. Bacteriophage from a cDNA library constructed in the vector lgt11 are plated under lytic growth conditions. After plaques appear, expression of the b-galactosidase fusion proteins encoded by the recombinant phage is induced by placing nitrocellulose filters impregnated with IPTG onto the plates. Phage growth is continued and is accompanied by the immobilization of proteins, from lysed cells, onto the nitrocellulose filters. The filters are lifted after this incubation, blocked with protein, then reacted with a radiolabeled recognition-site DNA (containing one or more binding sites for the relevant sequence-specific protein) in the presence of an excess of nonspecific competitor DNA. After the binding reaction, the filters are washed to remove nonspecifically bound probe and processed for autoradiography. Potentially positive clones detected in the primary screen are rescreened after a round of plaque purification. Recombinants which screen positively after enrichment and whose detection specifically requires the recognition-site probe (not detected with control probes lacking the recognition site for the relevant protein) are then isolated by further rounds of plaque purification.

COMMENTARY

Sequence-specific DNA-binding proteins function in diverse capacities including DNA replication, recombination, and transcription. In higher eukaryotic cells, biochemical approaches have been used to identify such proteins (see accompanying units in this chapter). Typically, these proteins bind specifically and with high affinity to distinct sequence motifs. Prior to the application of the cloning strategy described by Singh et al. (1988), genes encoding these proteins could only be isolated through purification of substantial amounts of the relevant activities (Weinberger et al., 1985; Kadonaga et al., 1987). The screening strategy discussed below obviates purification of a DNA-binding protein for the purpose of isolating its gene. Therefore, it is ideally suited for facilitating the analysis of rare regulatory molecules.

The strategy is designed to directly detect recombinant clones that encode sequence-specific DNA-binding proteins. It depends on the functional expression of the DNA-binding domain of a sequence-specific protein in E. coli at high levels and a strong interaction between this domain and its recognition site. If these conditions are fulfilled, a recombinant clone encoding a sequence-specific DNA-binding protein can be detected by probing nitrocellulose filter replicas of a lambda expression library with the corresponding recognition-site DNA (Singh et al., 1988; Staudt et al., 1988). This strategy is quite similar to that previously developed for the isolation of genes by screening recombinant expression libraries with antibodies specific for given proteins (Young and Davis, 1983;

see UNIT 6.7). In fact, the phage expression vector lgt11--designed for immunological screening--is also employed in the detection of DNA-binding protein clones. The simple procedure described in the basic protocol has permitted the isolation of many clones encoding different transcription factors including MBP-1 (Singh et al., 1988), Oct-2 (Staudt et al., 1988), E12 (Murre et al., 1989), XBP (Hsiou-Chi et al., 1988), IRF-1 (Miyamoto et al., 1988), and CREB (Hoeffler et al., 1988). The denaturation/renaturation cycle described in the alternate protocol may in some cases increase the detection signal by facilitating the correct folding of a larger fraction of the E. coli expressed protein. Alternatively, it may help to dissociate insoluble aggregates that form as a consequence of overexpression. This modified procedure has been successfully used to isolate clones encoding Oct-1 (Sturm et al., 1988), Oct-2 (Muller et al., 1988), Pit-1 (Ingraham et al., 1988), and RF-X (Reith et al., 1989).

Prospects for the isolation of cDNAs encoding sequence-specific DNA-binding proteins by this strategy can be assessed by examining the three assumptions on which it is based: (1) functional expression of the DNA-binding domain of the sequence-specific protein in E. coli; (2) a strong and specific interaction of the binding domain with its recognition site; and (3) high-level expression of the DNA-binding domain. A number of eukaryotic sequence-specific DNA-binding proteins have been functionally expressed in E. coli including the proteins GAL4 (Keegan et al., 1986) and GCN4 (Arndt and Fink, 1986) of yeast, engrailed of Drosophila (Desplan et al., 1985), E2 of the bovine papillomavirus (Androphy et al., 1987), and EBNA-1 of the Epstein-Barr virus (Rawlins et al., 1985). Thus it is reasonable to expect the functional expression of the DNA-binding domains of many eukaryotic sequence-specific proteins in E. coli. Interestingly, in most cases the functional DNA-binding domain is contained within a small tract of 100 to 300 amino acids. Therefore, successful screening is not dependent on full-length cDNA clones. It simply requires that a given recombinant expression library contain partial cDNA clones spanning the DNA binding domain of the sequence-specific protein. However, if this DNA-binding domain is encoded in the 5' region of a long mRNA, the frequency of functional cDNA clones in a library made by oligo(dT) priming is expected to be low. It is therefore preferable for the purpose of this screening strategy to construct a library by randomly priming cDNA synthesis (Staudt et al., 1988). Finally, it should be noted that some eukaryotic sequence-specific proteins have been shown to recognize target DNA only as heterodimers (Chodosh et al., 1988). Obviously, genes encoding the subunits of these proteins cannot be cloned by this strategy.

Successful screening may be restricted to proteins with relatively high binding constants since only these are likely to form complexes with half-lives long enough to withstand the wash protocol. For example, if the DNA-binding protein has an association constant of 1010 M-1, then under the screening conditions (the DNA probe in excess at a concentration of ~10-10 M), approximately half of the active molecules on the filter will have DNA bound. Since the filters are subsequently washed for 30 min, the fraction of protein-DNA complexes that remain will be determined by their dissociation rate constant. Assuming a diffusion-limited association rate constant of 107 M-1 S-1, the dissociation rate constant in solution will be 10-3 S-1. This rate constant translates into a half-life of ~10 min. Thus one-eighth of the protein-DNA complexes should survive the 30-min wash. For a binding constant of 109 M-1, about one-tenth of the active protein molecules will have DNA bound, but virtually all of this signal should be lost since the half-life of these complexes in solution is ~1 min.

However, it is unclear whether the equilibrium and kinetic constants of a protein-DNA interaction in solution accurately describe the reversible binding of a DNA probe to a matrix of protein immobilized on a filter. Thus, it may be possible to isolate recombinants encoding proteins with binding constants of <=109 M-1. These considerations nevertheless suggest that the sensitivity of this methodology for

low-affinity proteins might be significantly enhanced by using DNA probes containing multiple binding sites that are spaced such that the probe can simultaneously bind two or more immobilized protein molecules. Enhanced sensitivity with a multisite probe has been impressively demonstrated in the molecular cloning of a mammalian regulatory protein (Staudt et al., 1988). Finally, since the binding constants of sequence-specific proteins are dependent on ionic strength, temperature, and pH, manipulation of these factors might also enhance detection.

The DNA-binding domains of sequence-specific proteins need to be overexpressed in E. coli to permit detection with radiolabeled recognition-site probes. These proteins, when expressed to a level of ~1% of the total cellular protein, can be readily detected (Singh et al., 1988; Staudt et al., 1988). This level of recombinant protein expression is typical of lgt11.

Analysis of DNA-Protein Interactions Using Proteins Synthesized In Vitro from Cloned Genes

The availability of a cloned gene makes it possible to synthesize the encoded protein by in vitro transcription and translation. As described in UNIT 10.17, protein-coding sequences are cloned into a vector containing a promoter for SP6 or T7 RNA polymerase (UNIT 1.5), messenger RNA is produced by transcribing the DNA template (UNIT 3.8), and the desired protein is synthesized as a 35S-labeled species by in vitro translation. Such in vitro synthesized proteins are extremely useful for determining whether a cloned gene encodes a specific DNA-binding protein and for analyzing DNA-protein interactions. To detect DNA binding activity, the labeled protein is incubated with specific DNA fragments, and protein-DNA complexes are separated from free protein by electrophoresis in native acrylamide gels (UNIT 12.2). Unlike the more conventional mobility shift assay which utilizes 32P-labeled DNA and unlabeled protein, the assay described here generally utilizes 35S-labeled protein and unlabeled DNA. Major advantages of this method are that any desired mutant protein can be tested for its DNA-binding properties simply by altering the DNA template, and the subunit structure (e.g., dimer, tetramer) can be determined.

COMMENTARY

The procedure described here for analyzing DNA-protein interactions was first utilized in studies on the yeast GCN4 transcriptional activator protein (Hope and Struhl, 1985). It differs from conventional biochemical approaches in that the protein of interest is not obtained from cells, but rather is synthesized by in vitro transcription and translation of a cloned gene. However, once the protein is synthesized, many standard procedures for studying specific DNA-protein interactions (e.g., UNITS 12.2, 12.3, & 12.4) can be performed with only minor modifications. In addition, the method is extremely useful for analyzing the properties of mutant proteins and for determining subunit structure, issues that are much more difficult to investigate by more classical biochemical techniques.

The protocol for synthesizing [35S]protein by in vitro transcription and translation is detailed in UNIT 10.17. DNA-binding activity is detected by incubating the labeled protein with appropriate DNA fragments, and separating the protein-DNA complexes from free protein by electrophoresis in native acrylamide gels. This DNA-binding assay is essentially the reverse of the standard mobility shift assay