

Confirming the Functional Importance of a Protein–DNA Interaction

Important issues

- *The functional relevance of a protein–DNA interaction is difficult to establish.*
- *The hypothesis that an interaction is relevant can be tested by several different experiments, although none by itself is conclusive.*
- *The combined results of several experimental approaches are needed to rigorously examine the relevance of a protein–DNA interaction.*

INTRODUCTION	292
CONCEPTS AND STRATEGIES	294
Abundance of a protein–DNA complex in vitro	294
Relative expression patterns of the DNA-binding protein and target gene	295
Correlation between nucleotides required for protein binding and those required for activity of the control element	296
<i>trans</i> -Activation of a reporter gene or endogenous gene by overexpression of a DNA-binding protein	297
Cooperative binding and synergistic function of proteins bound to adjacent control elements	299
Comparison of genomic and in vitro footprinting patterns	301
Relative affinity of a protein–DNA interaction	302
Gene disruption or antisense experiments	304
Dominant-negative mutants	305
In vitro transcription strategies	308
In vivo protein–DNA crosslinking	310
Altered specificity experiments	313

INTRODUCTION

Discovering a gene encoding a novel DNA-binding protein has become a relatively straightforward task in the modern era of molecular biology, but definitively establishing that the protein directly regulates a target gene by binding to a defined control element is among the most difficult of tasks.

The preceding chapters described experimental strategies that lead to the identification of important *cis*-acting sequence elements, as well as proteins that bind those elements. In most instances, a protein identified has been implicated as a potential regulator of the gene of interest because it interacts with an important control element in an electrophoretic mobility shift assay (EMSA) or DNase I footprinting experiment. The identification of a specific DNA-binding protein provides a significant advance because it allows one to hypothesize that the protein is responsible for the function of the control element *in vivo*. However, by itself, the identification of a protein–DNA interaction does not demonstrate its relevance.

The detection of a protein–DNA interaction in a crude cell extract reflects many factors: (1) the abundance of the protein in the cells from which the extract was prepared, (2) the efficiency with which the protein was extracted from the cells, (3) the stability of the active protein within the extract, (4) the maintenance of essential posttranslational modifications during extract preparation, (5) the conditions used for the *in vitro* DNA-binding assay, and (6) the affinity of the protein for the isolated control element (Table 9.1).

The above criteria for detecting a protein–DNA interaction *in vitro* are very different from the criteria that determine which protein functionally interacts with the control element *in vivo* (i.e., which protein regulates the endogenous gene by binding to the control element of interest) (Table 9.1). These criteria include (1) the abundance and stability of the protein in the cell nucleus, (2) the affinity of the protein for the site, (3) the ability of the protein to carry out appropriate interactions with other proteins bound to adjacent sites and with non-DNA-binding cofactors, (4) appropriate posttranslational modifications that allow the protein to carry out the necessary protein–DNA and protein–protein interactions, and (5) the appropriate subnuclear localization of the protein.

When considering the above points, it is readily apparent that the detection of a protein–DNA interaction *in vitro* provides only weak evidence that it is relevant *in vivo*, even when the DNA sequence element to which the protein binds is known to be important (i.e., by mutagenesis). As discussed in Chapter 7, most DNA-binding proteins are capable of recognizing a wide range of DNA sequences with a wide range of affinities.

TABLE 9.1. *Factors that influence the in vitro detection and in vivo relevance of protein–DNA interactions*

Detection <i>in vitro</i>	Functional relevance <i>in vivo</i>
Abundance of protein in cell	Abundance and stability of protein in cell nucleus
Efficiency of protein extraction	Affinity of protein for the site
Stability of protein in extract	Ability of protein to carry out appropriate interactions with other proteins
Maintenance of essential post-translational modifications during extract preparation	Posttranslational modifications that allow the protein to carry out necessary interactions
Conditions used for <i>in vitro</i> DNA-binding assay	Appropriate subcellular and subnuclear localization of the protein
Affinity of the protein for the isolated control element	

Furthermore, most DNA-binding proteins are members of multiprotein families, with each cell type containing several family members that recognize similar DNA sequences. On the basis of these considerations, there is a high probability that multiple proteins will be capable of binding a defined control element *in vitro*, including several members of a particular protein family, and perhaps members of another family that recognize a similar or overlapping sequence. The difficult challenge is to determine which of these proteins is capable of carrying out the protein–protein and protein–DNA interactions that allow it to regulate the endogenous gene. If by chance only one predominant DNA-binding protein is detected *in vitro*, it may or may not be the one that is responsible for the function of the control element *in vivo*; other proteins within the cells will almost certainly be capable of binding the same element, even if they were not detected in the initial EMSA or DNase I footprinting studies.

A definitive approach for confirming the functional importance of a protein–DNA interaction in mammalian cells has not yet been developed. In fact, there are few examples of protein–DNA interactions that have been shown conclusively to be relevant. However, the situation is not as bleak as the above discussion might make it appear. Although the functional relevance of most protein–DNA interactions that have been reported in higher eukaryotes remains tenuous (including some of those studied in the laboratories of the authors), several have been shown to be relevant beyond reasonable doubt. In addition, it should be noted that relevant binding sites for some classes of transcription factors, such as ligand-inducible nuclear hormone receptors, are relatively easy to identify.

In the absence of a definitive experiment (i.e., an experiment that allows one to visualize directly a protein binding to a control element within its natural chromosomal location and regulating transcription of the linked gene), the only viable approach is to hypothesize that a protein–DNA interaction is relevant, and then to subject that hypothesis to as many rigorous tests as possible. This approach, of course, is not new, because it is a central tenet of the scientific method. If a number of independent experimental tests support the hypothesis, confidence that it might be correct is enhanced. The point at which a hypothesis of this sort has been confirmed beyond a reasonable doubt depends on a subjective evaluation and, therefore, is best left to the judgment of the scientific community.

In this chapter, twelve different approaches which can be used to test a hypothesis that a specific protein–DNA interaction is functionally important are presented. We describe the information gained from each approach and explain why each yields inconclusive results. The approaches vary widely with respect to the amount of effort required and the quality of information obtained. If a major goal is to establish the functional importance of an interaction, as many of these approaches as possible should be pursued.

Before proceeding, it is important to emphasize that a rigorous examination of the relevance of a protein–DNA interaction might not be an important objective of the laboratory. The primary interest may instead be to determine whether the DNA-binding protein that was discovered is important for a biological process. In this instance, the next step is more obvious: The gene encoding the protein can be disrupted in a cell or animal. If an interesting phenotype is observed, the laboratory would likely choose to begin a broadly based analysis of the DNA-binding protein, regardless of whether it carries out a functional interaction with the control element that originally was used for its identification. This is a common and valid course of action. However, if the long-term goal is to link a DNA-binding protein to its relevant target genes, or to carefully dissect the mechanism by which a target gene is induced by a constellation of DNA-binding proteins, cofactors, and general transcription factors, the issues discussed in this chapter will ultimately need to be considered.

CONCEPTS AND STRATEGIES

Abundance of a Protein–DNA Complex In Vitro

As stated above, the abundance of a protein–DNA complex in an in vitro binding assay, such as an EMSA or DNase I footprinting assay, is often used to develop an initial hypothesis that a given protein is responsible for the function of a control element. A predominant protein–DNA complex might be observed, for example, in an EMSA using a radiolabeled probe containing an important control element and a nuclear extract prepared from a relevant cell line or tissue. Although one can hypothesize that the interaction is relevant, the hypothesis is only weakly supported by this result for the reasons discussed in the introduction to this chapter.

The interleukin-12 (IL-12) promoter analysis discussed in Chapter 7 provides an example of this concept. To determine which protein is responsible for the function of the IL-12 promoter element between –88 and –99, an EMSA experiment was employed using extracts from uninduced and induced macrophages (Plevy et al. 1997). Several distinct protein–DNA complexes were detected. By adding specific antibodies to the binding reactions, the most abundant complexes were found to contain the C/EBP β protein. Less abundant complexes contained C/EBP δ and C/EBP α . These results led to the hypothesis that C/EBP β is responsible for the function of the –88/–99 element. However, the data do not rule out the possibility that, instead, C/EBP δ , C/EBP α , or another binding protein that does not yield a detectable EMSA complex is functionally relevant. C/EBP β may yield the most abundant complex because it (1) is the most abundant protein in the cell extracts, (2) binds to the isolated element with highest affinity, (3) is extracted from the cells more efficiently than other proteins, (4) is more stable in the extract than other proteins, (5) maintains posttranslational modifications that enhance DNA binding more effectively than other proteins, or (6) is more compatible with the particular EMSA conditions used than the other proteins. However, a different protein may be functionally relevant in vivo for any of a number of reasons, including the following: (1) A different protein may carry out optimal physical and functional interactions with other proteins that bind the promoter, (2) a protein that was less abundant than C/EBP β in the cell extract, and therefore undetectable, may bind the element with higher affinity in vivo, or (3) a protein that cannot be detected with the in vitro EMSA conditions used may preferentially bind the control element in vivo.

For the IL-12 promoter analysis, it will be particularly difficult to determine which C/EBP family member is functionally relevant. Although the most abundant EMSA complex contains C/EBP β , C/EBP δ binds the same sequence, and like C/EBP β , its abundance increases upon macrophage activation. The challenge of distinguishing between members of a multiprotein family is discussed frequently in this chapter.

Although multiprotein families contribute considerably to the challenge of establishing the functional relevance of a specific protein–DNA interaction, it can also be difficult to identify the relevant protein from among different families. The analysis of the terminal transferase (TdT) gene, which is expressed in immature lymphocytes, provides an example of this issue. DNase I footprinting studies led to the identification of a protein in nuclear extracts from immature lymphocytes that bound a critical control element in the TdT promoter, called the D' element (Lo et al. 1991). Upon purification and cloning, this protein was found to be Ikaros, suggesting that Ikaros or an Ikaros family member was the relevant activator of TdT transcription through the D' element (Georgopoulos et al. 1992; Hahm et al. 1994). However, subsequent experiments revealed that members of the Ets

family of transcription factors are capable of binding the D' element with considerable affinity, even though these proteins were not easily detected in EMSA or footprinting experiments with crude nuclear extracts (Ernst et al. 1993). By subjecting the Ikaros-D' and Ets-D' interactions to many of the tests described below, the relevance of the Ets-D' interaction for TdT activation was strongly supported, and the relevance of the Ikaros-D' interaction was not (Ernst et al. 1996). It remains unknown why Ikaros is the predominant D'-binding protein in nuclear extracts from TdT-expressing cells, despite several lines of evidence suggesting that it is not the relevant activator.

Relative Expression Patterns of the DNA-binding Protein and Target Gene

A second straightforward test of the hypothesis that a protein–DNA interaction is functionally relevant is a comparison of the expression patterns of the DNA-binding protein and putative target gene. This test can be carried out with transformed cell lines and primary cells by first quantifying steady-state mRNA levels or nascent transcripts (see Chapter 3) from the putative target gene. Target gene transcription can then be compared to the expression pattern of the proposed regulatory protein, which can be monitored by measuring steady-state protein levels (using immunoblot, immunoprecipitation, immunofluorescence, flow cytometry, or a biochemical assay).

The methods that are available for monitoring each DNA-binding protein and target gene vary. However, the preferred comparison would involve assays that are most meaningful for demonstrating a relationship between the DNA-binding protein and target gene; namely, a biochemical assay to quantify the DNA-binding activity (or capacity for transcriptional activation) of the protein, and a nuclear run-on assay to monitor nascent transcription of the target gene. If these assays are impractical because of the large number of cells needed to monitor DNA binding and the difficulty of the nuclear run-on, an alternative is to quantify the abundance of the DNA-binding protein and the steady-state mRNA abundance for the target gene.

The information gained by carefully comparing the expression patterns of the DNA-binding protein and target gene can either support or help to rule out a hypothesis. If the DNA-binding protein of interest is present in all of the cell types that express the target gene, the results would support the hypothesis that the protein is a relevant activator of the target. If, on the other hand, cells are identified that express the target gene, but not the putative activator, the hypothesis would be weakened. The hypothesis would not be negated, however, because some genes are activated by different sets of factors in different cell types (see, e.g., Lauring and Schlissel 1999).

It is important to note that a target gene is rarely, if ever, expressed in all of the cell types that express a relevant transcriptional activator. More likely, it is expressed in only a subset of the cell types expressing an activator. According to basic combinatorial principles of gene regulation (see Chapter 1), a given DNA-binding protein contributes to the transcription of many genes with varying expression patterns by acting in conjunction with several other transcription factors, each possessing its own unique expression pattern. In addition, numerous transcription factors are regulated by posttranscriptional mechanisms, allowing them to be present in an inactive state (or in an inappropriate subcellular compartment) in cells that do not express relevant target genes (e.g., NF- κ B, nuclear factor of activated T cells [NFAT], nuclear hormone receptors). For these reasons, a perfect correlation between activator and target gene is almost never observed; the most one can expect is the presence of the activator in all cells that express the target.

One additional limitation of the correlation between activator and target gene is that it does not provide a direct functional link between the two. The correlation may be fortuitous or the activator may indirectly regulate the target gene by regulating one of the gene's direct regulators.

For genes that are inducible in a cell line or tissue, the basic expression pattern analysis can be enhanced by a comparison of the kinetics of induction of the transcription factor and putative target gene. The cells can be treated with an appropriate inducing agent and the time course of induction of the DNA-binding activity in nuclear extracts can be compared to the time course of induction of the target gene (or a chromosomally integrated reporter gene). If the induction kinetics are similar or if the DNA-binding activity is induced slightly earlier than the target gene, the results would be consistent with the hypothesis that the protein regulates the gene. For some classes of transcription factors, such as nuclear hormone receptors, kinetic experiments can provide strong evidence that a protein–DNA interaction is relevant.

One caveat to the interpretation of kinetic experiments is that the precise concentration of a DNA-binding protein that needs to be present for a target gene to be activated is usually unknown. Perhaps the target gene can be activated when the DNA-binding protein reaches a concentration that is only 10% of its maximum concentration. If a target gene is transcribed when its putative activator is present at only 10% of its maximal concentration, the results at first glance would suggest that target gene induction is independent of the activator. In some instances, it has been found that a high threshold concentration of a critical DNA-binding protein must be present for target gene activation (Fiering et al. 1990). If the threshold concentration of a protein is indeed high, a kinetic analysis may be informative. However, in the absence of information about the threshold concentration, it is difficult to evaluate experiments that attempt to correlate the kinetics of induction of a transcription factor and target gene.

Correlation between Nucleotides Required for Protein Binding and Those Required for Activity of the Control Element

A third fundamental test of the hypothesis that a protein–DNA interaction is functionally relevant involves a detailed comparison of the nucleotides required for the function of the control element and the nucleotides required for binding of the putative transcription factor. This is an extremely powerful and underutilized method that can provide relatively strong support for the hypothesis that a particular DNA-binding protein, or at least a member of a particular family of DNA-binding proteins, is responsible for the activity of the control element.

The TdT promoter analysis provides an example of this strategy. As mentioned previously, the important D' element can interact with both Ikaros and Ets proteins. Ikaros was the predominant D'-binding protein observed in extracts from TdT-expressing cells, leading to the hypothesis that it might be the functional activator. The expression patterns of Ikaros and Ets proteins did not help to determine which protein was the functional activator, because proteins of both families were expressed in all TdT-expressing cells examined. A detailed mutant analysis of the D' element appears to have been a useful strategy for identifying the relevant protein family (Table 9.2) (Ernst et al. 1996). A series of single- and double-base substitutions in the TdT D' element were constructed and tested for their effect on promoter function in transient and stable transfection experiments, and for their effect on the binding of Ikaros and Ets proteins. As summarized in Table 9.2, the results revealed that the nucleotides required for Ikaros binding were significantly different from the nucleotides required for promoter activity in both transfection assays. Two different

TABLE 9.2. Effect of substitution mutations on promoter function and protein binding to the TdT D' element

D' Sequence	Promoter function	Ets binding	Ikaros binding
WT G C A G G A A G T T G	+	+	+
m83 G <u>A</u> A G G A A G T T G	+/-	+/-	+
m84 G C C G G A A G T T G	+++	++	-
m85 G C A <u>I</u> I A A G T T G	-	-	-
m86 G C A G G C C G T T G	-	-	-
m87 G C A G G A A <u>I</u> T T G	-	-	+
m88 A C C G G A A G T <u>A</u> G	+++	++	-

mutations enhanced promoter activity while abolishing Ikaros binding, and another mutation abolished promoter activity without affecting Ikaros binding. In contrast, the nucleotides required for promoter activity matched precisely the nucleotides required for the binding of various proteins of the Ets family.

The above results support the hypothesis that an Ets family protein is a functional activator of TdT transcription. However, this strategy has significant limitations. In particular, although the strategy can be useful for addressing the relevance of different families of proteins, it usually will provide little distinction between members of the same family. A second limitation is that the assays used to monitor the effect of mutations, such as the transfection assays, are artificial (see Chapter 7). In the TdT promoter analysis, the nucleotide requirements for activity of the D' element in the transfection assays are likely to reflect the requirements for D' activity in the endogenous gene. However, the high plasmid copy number during the transfection assay and the removal of the control region from its natural context have the potential to alter the outcome of the analysis. A final limitation is that careful mutant studies of this type can be much more difficult to evaluate if the function of the element depends on the simultaneous binding of two or more proteins. In this instance, it may be very difficult to demonstrate a correlation between the nucleotides required for activity and the nucleotides required for the binding of an individual protein. Despite these limitations, this strategy often provides a useful test of a hypothesis.

***trans*-Activation of a Reporter Gene or Endogenous Gene by Overexpression of a DNA-binding Protein**

The ability of an overexpressed or ectopically expressed protein to *trans*-activate a reporter gene regulated by the control region of interest, or to *trans*-activate an endogenous gene, can provide support for the functional relevance of a protein–DNA interaction. However, these experiments are difficult to interpret when the DNA-binding protein is expressed at a higher concentration than is found in a normal cell. The presence of multiple copies of a reporter plasmid in a transfected cell can lead to similar interpretation problems.

An experiment that is commonly performed begins with the insertion of a cDNA encoding the DNA-binding protein into a vector that drives expression following introduction into cultured cells (Fig. 9.1). For mammalian cells, common expression vectors contain a strong viral promoter/enhancer, such as that derived from cytomegalovirus. Cultured cells are then cotransfected with this expression plasmid and a reporter plasmid regulated by the control region of interest; the reporter assay is used to monitor the effect

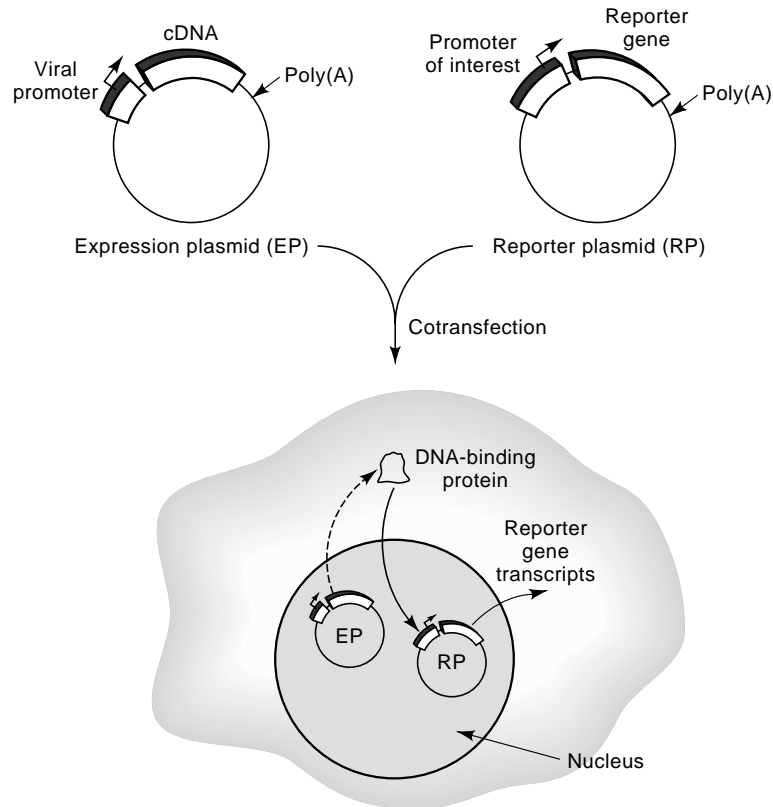


FIGURE 9.1. *trans*-Activation of reporter gene expression by ectopic overexpression of a transcription factor. (From *Molecular Cell Biology* by Lodish et al. Copyright 1986, 1990, 1996, by Scientific American Books, Inc. Used, with permission, by W.H. Freeman and Company.)

of the overexpressed protein on the activity of the control region. If overexpression results in activation of the control region, the requirement for the protein's binding site can be assessed by repeating the experiment with a reporter plasmid containing a binding site mutant. In this experiment, *trans*-activation should not be observed.

A positive result with this type of experiment strongly suggests that the DNA-binding protein can activate the control region *when both the DNA-binding protein and control region are overexpressed*. However, the result provides little evidence that the protein, when expressed at physiological concentrations, can regulate the two copies of the endogenous target gene present in diploid cells. Current models suggest that genes are regulated by multiple protein–DNA and protein–protein interactions (see Chapter 1). By substantially increasing the concentration of a protein that is not normally involved in regulating a gene, aberrant protein–protein and/or protein–DNA interactions might take place that are sufficient for the gene to be activated or repressed. Overexpression of the reporter plasmid containing the control region of interest can also enhance protein–DNA interactions that do not normally occur.

One example of this experimental approach is provided by the IL-12 promoter analysis (Plevy et al. 1997). To test the hypothesis that C/EBP β is a relevant activator through the –88/–99 element, uninduced macrophages were cotransfected with a C/EBP β expression plasmid and an IL-12 promoter-reporter plasmid. Overexpressed C/EBP β enhanced promoter activity in uninduced cells to a level comparable to that observed in induced cells in the absence of overexpression. Mutation of the C/EBP β -binding site abolished the *trans*-

activation. At first glance, these results appear to suggest that C/EBP β is a relevant activator of the IL-12 promoter. However, they merely show that the promoter can be activated when C/EBP β and the IL-12 promoter-reporter plasmid are present at unusually high concentrations. The results provide no significant evidence that C/EBP β , when present at physiological concentrations, is a relevant activator of the endogenous IL-12 p40 gene.

These same issues must be considered when an overexpressed transcription factor is found to *trans*-activate an endogenous gene. However, in this instance, a positive result may provide a modest level of support for the hypothesis being tested. First, if the overexpressed DNA-binding protein is capable of *trans*-activating the endogenous gene in a cell type that does not normally express the gene, it must be capable of carrying out protein–DNA and protein–protein interactions that are of sufficient specificity and affinity to overcome chromatin barriers, etc. Second, because only two copies of the endogenous gene exist in a diploid cell, strong activation by a protein that is not physiologically relevant seems less likely; to achieve significant levels of transcription from an endogenous gene, highly specific interactions with other proteins may be required.

One potential solution to the overexpression problem is to express the protein ectopically at a concentration comparable to that found in normal cells. This might be possible by stably expressing the protein. Individual cell clones can then be isolated and tested for protein expression level. Clones that express the protein at a level comparable to that found in a cell line that naturally expresses it can be used for further analysis of target gene transcription. Although this strategy may lead to informative results, it may fail; ectopic expression of a single DNA-binding protein at a normal concentration is unlikely to be sufficient to activate a target gene, unless it is truly the only regulator of that gene missing from the cells being used. If other tissue-specific proteins are needed, the gene will not be efficiently activated.

A second set of experiments that can lessen the concern about protein overexpression is to compare the activities of several members of a transcription factor family when each is overexpressed to a similar extent. If only one family member can *trans*-activate a control region, it must bind the control element with an unusually high affinity, or it must be capable of carrying out specific interactions with other proteins needed for the control region to function. In the IL-12 promoter analysis, for example, Murphy et al. (1995) determined the extent of *trans*-activation following overexpression of various combinations of Rel proteins. The simultaneous overexpression of p50 and c-Rel resulted in much stronger *trans*-activation than overexpression of other Rel proteins, either alone or in other combinations. An important control for this experiment is to show that all of the proteins are expressed at similar levels and are similarly active with a reporter plasmid that does not exhibit a preference for p50/c-Rel binding. Although these experiments used overexpressed proteins, the specificity observed suggests that a p50/c-Rel heterodimer may indeed be a functional activator of IL-12 transcription.

Cooperative Binding and Synergistic Function of Proteins Bound to Adjacent Control Elements

Support for the hypothesis that a protein–DNA interaction is functionally relevant can sometimes be provided by the selective ability of the protein to bind cooperatively with other proteins that interact with the control region. Support can also be provided by the selective ability of the protein to synergize functionally with other proteins (functional synergy does not necessarily involve cooperative binding; see Chapter 1).

The specific interaction between PU.1 and Pip at the Ig κ 3' enhancer provides an example of how a hypothesis can be supported by cooperative binding (Pongubala et al. 1992; Eisenbeis et al. 1995). Two of the critical elements within this enhancer are immediately adjacent to one another and appear to act in synergy. EMSA experiments performed with nuclear extracts and a radiolabeled probe spanning the two elements yielded a complex containing proteins bound to both sites (Fig. 9.2, lane 1) (Pongubala et al. 1992). DNA binding by the two proteins was highly cooperative as mutation of either site strongly reduced protein binding to the probe (Fig. 9.2, lanes 4–8). One protein within the complex was found to be an Ets protein named PU.1, and the other was found to be a member of the IRF family, named Pip or NF-EM5 (Eisenbeis et al. 1995).

The detection of a stable complex containing proteins bound cooperatively to two functionally important sites provides strong evidence a priori that these proteins (and not other members of the Ets and IRF families) are the relevant regulators of the Ig- κ enhancer. As discussed earlier, the detection of an EMSA complex with an isolated control element provides relatively weak evidence that the protein is functionally relevant because the criteria for the *in vitro* detection of a protein–DNA interaction are very different from the criteria for a relevant *in vivo* interaction. Nevertheless, if an EMSA complex observed with a crude extract contains proteins bound cooperatively to two sites that are functionally synergistic *in vivo*, the probability that the proteins within the complex are functionally relevant is substantially increased.

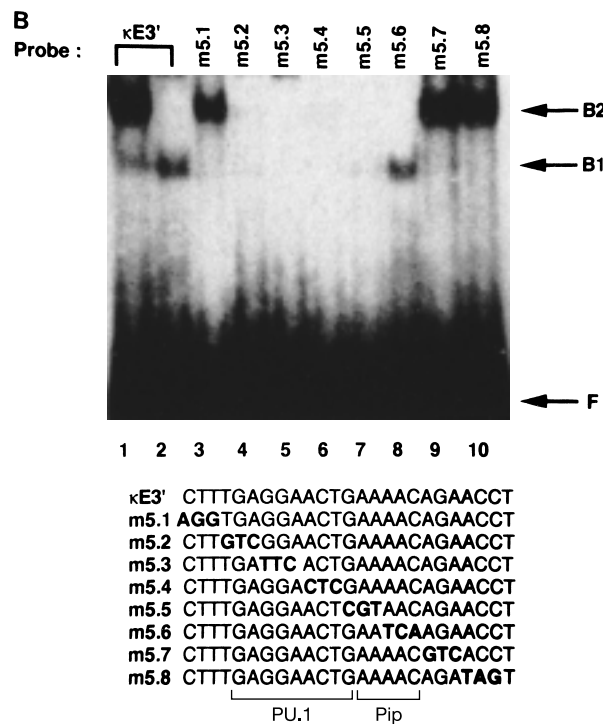


FIGURE 9.2. Highly cooperative binding of PU.1 and Pip to the Ig κ 3' enhancer. EMSA experiments were performed with radiolabeled oligonucleotides containing the wild-type or mutant enhancer sequences shown at the bottom. Binding reactions in lanes 1 and 3–10 were performed with nuclear extracts from the S194 cell line. The binding reaction in lane 2 contains *in vitro* translated PU.1. Band B2 contains PU.1 and Pip cooperatively bound to the probe. Band B1 contains PU.1 alone. Band F corresponds to free probe. The locations of the PU.1 and Pip-binding sites are depicted at the bottom. (Adapted, with permission, from Pongubala et al. 1992 [Copyright American Society for Microbiology].)

A key limitation of this assay is that strong, cooperative binding and functional synergy are observed with only a small subset of protein–DNA interactions. If cooperative binding to the control region is not observed *in vitro*, this strategy will not be useful. The PU.1/Pip example is particularly powerful because the cooperatively bound proteins are easily detected in EMSA experiments using crude nuclear extracts. If cooperative binding was observed only when the two recombinant proteins were added to an EMSA experiment, the results would provide much weaker support for the hypothesis. In this case, a demonstration of the selectivity of the cooperative binding would be needed, by comparing various Ets and IRF family members.

A second example of the cooperative binding and synergistic activation strategy, which highlights its limitations, recently emerged from studies of the interferon- β (IFN- β) enhancer by the Maniatis laboratory (Wathelet et al. 1998). Previous studies had found that an IRF family member, IRF-1, interacts with a functionally important enhancer element and binds cooperatively to the enhancer with other transcription factors, including ATF-2/c-jun and NF- κ B (Thanos and Maniatis 1995; Kim and Maniatis 1997). IRF-1 also can *trans*-activate the enhancer in synergy with those factors. The ability of IRF-1 to carry out cooperative and synergistic interactions at the enhancer suggested that it was a relevant activator of IFN- β transcription. The recent study by Wathelet et al. (1998), however, appears to have disproven that hypothesis. In this study, a protein complex was found to interact with the IRF recognition element following induction of IFN- β transcription. The inducible protein complex did not contain IRF-1, but instead contained two other IRF family members, IRF-3 and IRF-7. Several additional experiments confirmed that IRF-3 and IRF-7, but not IRF-1, contribute to the activation of IFN- β transcription *in vivo*. In particular, a single-base-pair substitution in the enhancer was identified that abolishes IRF-1 binding, but has no effect on IRF-3/IRF-7 binding or on enhancer activity. Thus, despite the initial evidence that IRF-1 can bind cooperatively to the enhancer with other relevant factors, and synergistically activate the enhancer in concert with those factors, it appears to be irrelevant for enhancer activity or IFN- β transcription.

The IRF-1 studies were misleading primarily because recombinant proteins were initially used to demonstrate cooperative binding and synergistic activation. This contrasts with the PU.1/Pip studies where the cooperative interactions were first observed in crude nuclear extracts from cells that express the Ig κ gene; recombinant proteins and overexpressed proteins were not needed. When cooperative interactions were finally observed with the IFN- β enhancer in crude nuclear extracts from induced cells (in the absence of protein overexpression), the complex was found to contain IRF-3 and IRF-7 instead of IRF-1.

Comparison of Genomic and In Vitro Footprinting Patterns

A straightforward test of the hypothesis that a specific protein–DNA interaction is relevant is to compare the footprinting pattern observed *in vitro* with a known DNA-binding protein to the pattern observed in genomic footprinting experiments with cells that actively express the putative target gene. This comparison can be carried out using any of the genomic footprinting methods described in Chapter 10, including DNase I footprinting or dimethyl sulfate (DMS) protection. If the footprinting pattern observed *in vitro* matches the pattern observed *in vivo*, the protein analyzed *in vitro* might actually be the protein bound to the endogenous gene when it is active, strongly suggesting that it is the relevant activator. If DNase I footprinting is used to perform this comparison, it usually will be necessary to use ligation-mediated PCR (LM-PCR) technology (see Chapter 10) for both the *in vitro* and genomic footprinting procedures, because it is difficult to compare footprint-

ing results obtained by LM-PCR with those obtained using radiolabeled probes. In contrast, if DMS footprinting is used, the results obtained by LM-PCR *in vivo* can be compared directly to the results obtained *in vitro* using a radiolabeled probe.

This strategy was used during the analysis of the lymphocyte-specific RAG-2 promoter (Lauring and Schlissel 1999). *In vitro* EMSA studies demonstrated that the B-cell activator, BSAP, can interact with an important control element. These results suggested that BSAP is an important activator of RAG-2 transcription in B cells. An *in vivo* interaction at the element was observed by DMS genomic footprinting, but *in vivo* footprinting results do not reveal the identity of an interacting protein. To provide evidence that the protein bound to the site *in vivo* was BSAP, a DMS protection analysis was performed *in vitro* using recombinant BSAP. The *in vivo* and *in vitro* interaction patterns were identical, suggesting that BSAP is indeed the protein that occupies the endogenous element in B cells.

The primary limitation of this strategy is the following: Even if the *in vitro* and genomic footprinting patterns are identical, the results do not usually distinguish between the various members of a protein family, because most family members lead to a similar footprint. Furthermore, some DNA-binding proteins do not yield footprinting patterns that are sufficiently unique, making it difficult to determine whether the protein bound to the endogenous locus *in vivo* is related to the known protein being tested *in vitro*.

Nevertheless, if clear genomic and *in vitro* footprints with characteristic properties are obtained, this comparison can provide compelling support for the relevance of a particular family of proteins for the function of a defined control element.

Relative Affinity of a Protein–DNA Interaction

One property of a protein–DNA interaction that may support its relevance is its affinity, relative to the affinities of other protein–DNA interactions that can occur at the same site. To our knowledge, the relationship between affinity and functional relevance of a transcription factor has not been examined carefully and systematically (although, in model systems, the affinity of a transcription factor for its target site is proportional to activation; see, e.g., Mauxion et al. 1991; Lehman et al. 1998). In some cases, such as the PU.1/Pip example described above, low-affinity interactions by the individual DNA-binding proteins are preferred so that the activity of the composite element remains dependent on cooperative binding by two or more proteins. Nevertheless, it seems possible that in many instances, the protein that functionally interacts with a control element will bind with an unusually high affinity, relative to the affinities of irrelevant proteins capable of binding the same site.

Affinity may be a particularly useful criterion for determining which member of a multiprotein family is most likely to carry out a functional interaction with an element. This assertion is based on the notion that the various members of a multiprotein family, although recognizing very similar sequences, have subtle sequence preferences. These sequence preferences may play a major role in determining which factor acts on a given control element. If this hypothesis is correct, a careful comparison of binding affinities of the members of a multiprotein family for a site might provide insight into the functionally important family member.

To compare relative affinities of a variety of proteins for a given site, several methods can be considered. One method is to obtain the pure recombinant proteins to be compared, and then to carry out careful K_d measurements as described in Chapter 13. This method will provide valuable information, but it has two disadvantages. First, it is difficult to prepare and accurately quantify recombinant forms of all the proteins that should be tested, in particular when working with a large multiprotein family. More importantly, the affinity of a recombinant protein for a DNA sequence may be significantly different from

that of the native protein. One reason for this potential difference is that protein–DNA interaction affinities can be altered by posttranslational modifications or by interactions with other proteins. In particular, a growing number of proteins are being found to possess autoinhibitory domains that reduce their affinities for DNA unless appropriately modified (see, e.g., Jonsen et al. 1996).

Thus, a more informative experiment would be to compare relative affinities in a cell extract, where the proteins are more likely to exist in their native state. If most of the proteins of interest within an extract are sufficiently concentrated to permit their detection in a basic EMSA, a rough comparison of their affinities for a site can be obtained. One approach is to determine the susceptibility of each protein–DNA complex to nonspecific competitor DNA, salt, or detergent by titrating each into the binding reactions. Higher-affinity interactions are likely to be more resistant to these reagents, particularly if one is comparing members of a family. A limitation of this approach, however, is that it will result in a comparison only of those proteins that are sufficiently concentrated to detect in a basic EMSA experiment. Proteins that are of relatively low abundance, or whose binding properties are incompatible with the EMSA conditions being used, cannot be tested.

An alternative approach was carried out in one of our laboratories with the TdT D' element discussed above (Fig. 9.3) (Ernst et al. 1996). The goal was to determine which Ets protein within an extract from a TdT-expressing cell line binds to the D' element with highest affinity, to identify a candidate for the functional activator of TdT transcription. Several known, and perhaps other novel, Ets proteins are present in TdT-expressing cells, but only a few EMSA complexes containing Ets proteins can be detected using nuclear extracts. To compare the relative affinities of proteins within the extract for the D' element, in the absence of a bias toward known proteins or proteins that could be detected by EMSA using crude extracts, sequence-specific DNA affinity chromatography was employed (see Chapter 8). The expectation was that the highest-affinity proteins would elute from an affinity column with the highest salt concentration. Indeed, following affinity chromatography, a single abundant EMSA complex was observed in the high-salt eluates, with all other EMSA complexes more abundant in the lower-salt eluates. A silver-stained protein gel led to the identification of the protein responsible for the complex (complex Y), and peptide sequencing revealed that the protein was an Ets protein called Elf-1. Immunoblot experiment of column fractions confirmed that Elf-1 eluted at higher salt concentrations than several other Ets family members. Subsequent experiments, like those described in this chapter, provided additional support for the hypothesis that Elf-1 is a functional activator of TdT transcription through the D' element. However, the hypothesis remains unproven. Thus, it is not yet known whether affinity provides a valid criterion for assessing functional relevance.

The affinity chromatography strategy possesses several advantages relative to the other two strategies mentioned above. First, the native proteins in the crude extract may retain posttranslational modifications that were present in the intact cells. Second, this method is not biased toward the most abundant proteins or previously described proteins. In fact, this strategy should result in the purification of any protein of reasonable abundance that binds the site with high affinity. Finally, if the high-affinity binding protein is novel, it can be identified or cloned by microsequencing (see Chapter 8).

A notable limitation of this approach is that the posttranslational modifications present in the intact cell might not be retained in the extract. Furthermore, the use of affinity chromatography as a measure of relative affinities might be somewhat inaccurate if the different proteins that bind a site rely to different extents on electrostatic interactions between the DNA-binding domain and DNA. If the electrostatic interactions are substantially different, salt elution from an affinity column will not provide an accurate assessment of relative affinities.

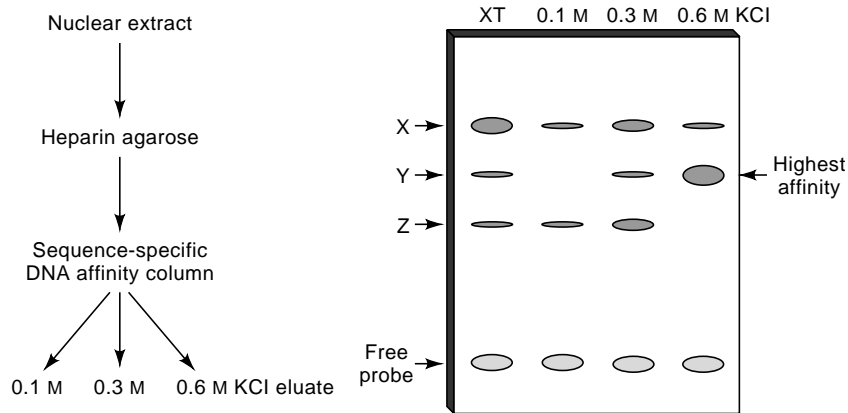


FIGURE 9.3. Comparison of protein–DNA interaction affinities in crude nuclear extracts by sequence-specific DNA affinity chromatography. Nuclear extracts from the murine T-cell line RLm11 were initially fractionated by heparin agarose to remove nucleases. Sequence-specific DNA affinity chromatography was then performed using a resin with covalently linked oligonucleotides containing the D' element from the TdT promoter. The sample was applied to the column in a buffer containing 0.1 M KCl. Proteins were then eluted with 0.3 and 0.6 M KCl. At the right, an idealized EMSA result is depicted; three protein–DNA complexes were detected when the extract was used (lane 1), but only one (complex Y) was preferentially retained on the column until the 0.6 M KCl elution (lane 4).

Gene Disruption or Antisense Experiments

A DNA-binding protein can be linked to a putative target gene by abolishing expression of the protein in a cell line or animal, using homologous recombination or antisense RNA technology (Capecchi 1989; Branch 1998; Stein 1998). The absence of the DNA-binding protein might result in a reduction of target gene expression, implicating the protein as a regulator of the gene.

Although gene disruption via homologous recombination and antisense have proven to be invaluable for assessing the importance of DNA-binding proteins for specific biological processes, the information provided about primary target genes is much less compelling. The phenotype of a mouse or cell line that lacks expression of a DNA-binding protein undoubtedly results from the altered expression of one or more genes. However, in most instances, it is difficult to determine which genes are directly responsible for the phenotype and which are direct targets of the DNA-binding protein. If the expression of a particular gene is diminished in a mutant animal, the gene could be a direct or indirect target of the DNA-binding protein. In other words, the transcription factor may directly regulate the gene, or it may regulate the expression of other genes that influence the expression of the gene of interest. If the absence of the DNA-binding protein does not disrupt the expression of a gene of interest, the gene may still be a direct target of the protein. Functional redundancy between two DNA-binding proteins might obscure their functions. Alternatively, a related DNA-binding protein might be up-regulated, allowing it to compensate for the disrupted protein.

To provide an example of these concepts, we can return to the TdT promoter and its potential regulation by Ikaros. As described above, Ikaros binds with high affinity to the critical D' element within the TdT promoter. On the basis of this result, Ikaros was originally considered a likely activator of TdT transcription. Support for this hypothesis was provided by an Ikaros gene disruption experiment (Georgopoulos et al. 1994). The mutant mice lack all cells of the B- and T-lymphocyte lineages, including the earliest progenitors of both lineages. Because TdT is one of the first genes activated when a hematopoietic stem cell becomes committed to the B- and T-cell lineages, the phenotype supported the

hypothesis that TdT is a functionally relevant target of Ikaros: (1) Ikaros binds to a critical control element in the TdT promoter, (2) the TdT gene is activated in progenitor B and T cells, and (3) the Ikaros mutant mice specifically lack the progenitor B and T cells that express the TdT gene.

Despite the fact that this model was attractive and supported by the available data, the experiments described earlier in this chapter suggest that Ikaros is not an activator of TdT transcription. In fact, recent data suggest that Ikaros is actually a repressor of TdT transcription at a later stage of lymphocyte development. Thus, the absence of TdT expression in Ikaros mutant mice is likely to be an indirect effect.

An example of the effect of transcription factor redundancy on the interpretation of gene disruption results was provided by studies of the LEF-1 and TCF-1 proteins implicated in regulation of the T-cell receptor α (TCR α) enhancer (Clevers and Grosschedl 1996; Okamura et al. 1998). LEF-1 and TCF-1 are highly homologous members of the HMG-1 family of proteins. Both proteins can bind a critical element within the TCR α enhancer and can *trans*-activate the enhancer in concert with DNA-binding proteins that bind adjacent elements, with LEF-1 *trans*-activation approximately 10 times stronger than TCF-1 *trans*-activation. Disruption of either the LEF-1 or TCF-1 genes revealed little effect on TCR α gene transcription, however (van Genderen et al. 1994; Verbeek et al. 1995). If only one of these proteins were known to exist, it would have been tempting to speculate, on the basis of the gene disruption result, that the protein was not a relevant activator of the TCR α enhancer. Because both proteins had been discovered, however, the possibility of redundancy was considered and examined by generating mice in which both genes were disrupted. The LEF-1^{-/-}TCF-1^{-/-} mice exhibited a severe defect in thymocyte development and were deficient for TCR α gene transcription, strongly suggesting that the two proteins are at least partially redundant.

These results highlight the fact that gene disruption and antisense phenotypes must be interpreted cautiously, and that other approaches are needed to evaluate rigorously the relevance of a DNA-binding protein for the regulation of a putative target gene.

Dominant-negative Mutants

By definition, a dominant-negative mutant of a protein is a protein variant that, when expressed in a cell containing the wild-type protein, disrupts the functions of that protein. Disruption usually occurs because the mutant retains some, but not all, of the wild-type protein's activities, allowing it to compete, albeit nonproductively, for an important target or substrate of the wild-type protein. Dominant-negative mutants of a DNA-binding protein can sometimes provide information about the functional importance of a protein–DNA interaction. However, the information provided by these experiments is limited, and the type of dominant negative used has a notable impact on the interpretation of the data.

A common method for performing dominant-negative studies is to cotransfect cells with a reporter plasmid containing a standard reporter gene regulated by the control region of interest and an expression plasmid for a dominant-negative version of the DNA-binding protein of interest (Fig. 9.4). The dominant-negative protein that is expressed is often designed to retain the capacity to bind DNA but not to carry out transcriptional activation functions. In other words, the protein contains its DNA-binding domain and lacks its transcriptional activation domains (see Chapter 12). The expectation is that this mutant protein will compete with the wild-type endogenous protein, thereby preventing the wild-type protein from activating reporter gene transcription (Fig. 9.4). Inhibition of transcription with this type of dominant-negative protein is often used to support the hypothesis that the wild-type protein is a functional activator of the control region fused to the reporter.

As a recent example, this approach was used by Plevy et al. (1997) to support the hypothesis that C/EBP β is a relevant activator of the IL-12 p40 promoter in lipopolysaccharide (LPS)-activated macrophages. Increasing concentrations of an expression plasmid for a dominant-negative form of C/EBP β , called LIP (Descombes and Schibler 1991), were cotransfected into the macrophages along with IL-12/CAT reporter plasmid. LIP contains the bZIP DNA-binding and dimerization domains of C/EBP β , but lacks its transcriptional activation domains. The cells were then activated with LPS, and the effect of LIP on IL-12 promoter activity was monitored using the CAT assay. The results showed that promoter activity was suppressed by LIP. Important control experiments showed that LIP expression had no effect on reporter plasmids containing promoters lacking C/EBP-binding sites (e.g., a CMV-CAT reporter).

Dominant-negative experiments of this type can support a hypothesis regarding the functional relevance of a DNA-binding protein, such as C/EBP β . However, a careful con-

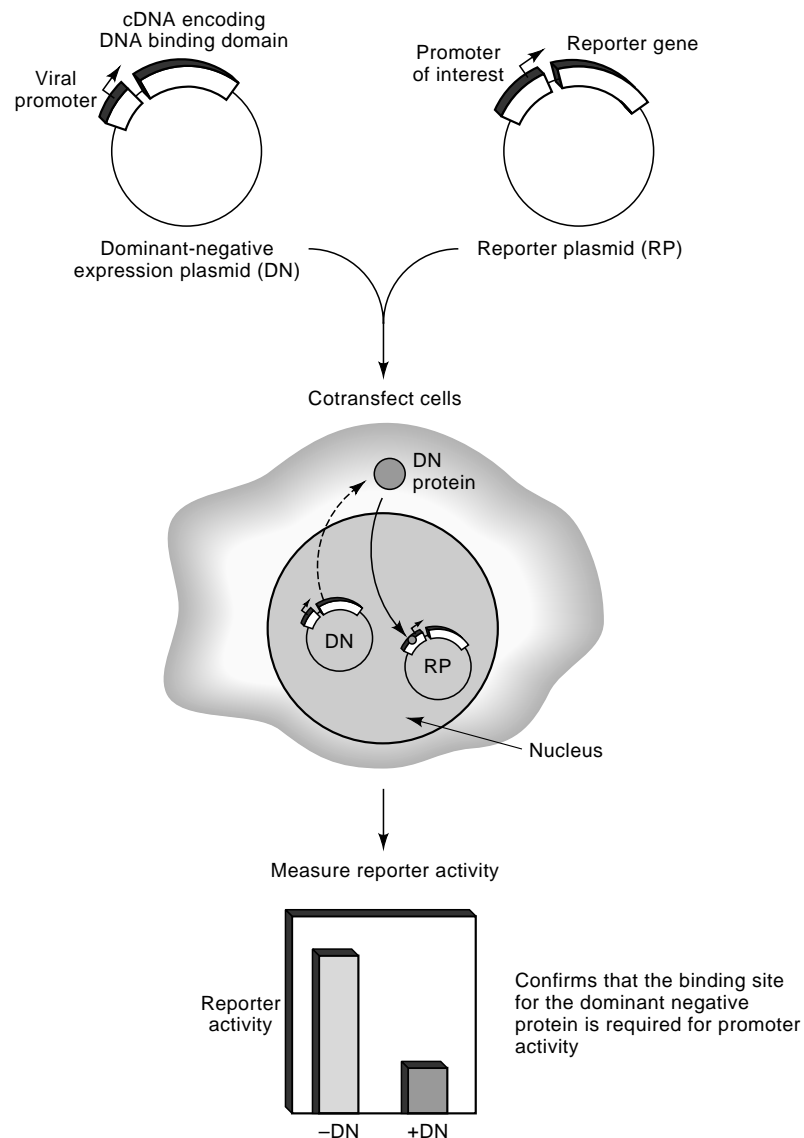


FIGURE 9.4. Inhibition of reporter activity by a dominant-negative version of a transcription factor that retains its DNA-binding domain.

sideration of the experiment reveals that the result merely confirms that the DNA-binding domain, when overexpressed, is capable of binding to the important control element, where it blocks activation by the functional activator. Regardless of the identity of the functional activator, the overexpressed dominant-negative mutant may block the important protein–DNA interaction, simply by occupying the binding site. Thus, the result provides two relatively modest pieces of information: (1) It demonstrates that the control element to which the dominant-negative protein binds is important for function of the control region, a result already established by the promoter mutant analysis; and (2) it demonstrates that the DNA-binding domain within the dominant-negative protein is capable of binding to that control element *in vivo* when overexpressed.

An alternative experiment is to test the effect of the same dominant-negative mutant on expression of the endogenous gene. This experiment provides an opportunity to confirm that the control element found to be important in an artificial transfection assay is important for regulation of the endogenous gene. However, this experiment suffers from most of the above limitations and from two additional caveats. First, it is difficult to rule out the possibility that the dominant-negative protein indirectly inhibits transcription from the endogenous target gene by altering expression of other genes within the cell. For example, the dominant-negative protein may inhibit expression of cellular genes that are needed for survival, leading to toxicity that might indirectly inhibit transcription from the target gene of interest. (The possibility of indirect effects must also be considered when analyzing the effect of a dominant negative on a transfected reporter plasmid, but this possibility can be addressed, at least in part, by testing a mutant reporter.) The second limitation of analyzing an endogenous target gene is that it is technically more difficult than analysis of a transiently cotransfected dominant-negative mutant and reporter plasmid. In the cotransfection assay, virtually every cell that takes up the reporter plasmid takes up the dominant-negative expression plasmid. Thus, the cell population can be analyzed for an effect of the dominant-negative protein on reporter gene transcription. In contrast, to monitor the effect of the dominant-negative protein on endogenous gene transcription, one must keep in mind that only a fraction of the cells are likely to be transfected and express the dominant-negative protein. Because the majority of cells will not be transfected, it may be difficult to observe the effect of the dominant negative on target gene expression.

Despite this added difficulty, there are several fairly straightforward solutions. First, the dominant-negative protein can be expressed from a high-titer retroviral vector that can infect all of the cells (see Chapter 11), perhaps leading to a measurable effect on transcription of the putative target gene. Second, the dominant-negative protein can be expressed in the cells by stable transfection. Preferably, the dominant-negative protein should be under the control of an inducible promoter, so that any toxicity of the protein is not manifested during the selection process (see Chapter 11). Third, the cells that are transiently transfected with the dominant-negative expression plasmid could be distinguished from the untransfected cells by flow cytometry or a related immunologic method (see Chapter 5, Box 5.5).

Although the dominant-negative proteins described above have limited value, other types of dominant-negative proteins may be more useful for specific types of DNA-binding protein families. In particular, dominant-negative mutants can be used to monitor the importance of proteins that bind DNA as dimers, such as b-ZIP and bHLH proteins. For these classes of proteins, dominant-negative versions of the protein can be expressed that retain the dimerization domain but lack the DNA-binding domain. If these dominant-negative proteins inhibit transcription from the putative target gene, the result would suggest that a protein capable of dimerizing with the dominant-negative protein is the functional activator through the control element of interest. Depending on the nature of the dimer-

ization domain, this result may implicate an entire family of DNA-binding proteins, or it may suggest that a protein among a specific subset of family members is likely to play a role. To confirm the relevance of this result, it may be necessary to test the dimerization domain's capacity to interact with various members of the protein family.

A classic example of a dominant-negative protein that acts by dimerizing with specific transcription factors is the HLH protein, Id1 (Benezra et al. 1990). Id1 and other Id family members contain helix-loop-helix domains that allow them to form heterodimers with bHLH activators, but they lack the basic region needed for DNA binding. The Id proteins therefore inhibit transcriptional activation by bHLH proteins by preventing them from binding to DNA. Thus, Id proteins can be used to assess the relevance of the bHLH family of proteins for the function of a control element. However, because each Id protein can dimerize with multiple bHLH proteins (Langlands et al. 1997), an inhibition of transcriptional activation provides limited insight into the identity of the functionally relevant family member.

In Vitro Transcription Strategies

In vitro transcription experiments can be used to support the relevance of a protein-DNA interaction. The basic approach begins with the development of an in vitro transcription assay for the control region, which can often be quite challenging (see Chapter 5 for cited examples of regulated transcription in nuclear extracts, and Chapter 14 for in vitro transcription methodology). Nucleosome reconstitution may aid in the assay's development (see Chapters 5 and 14). To determine whether the DNA element of interest contributes to the in vitro activity, mutations in the element can be tested. If the assay is found to be dependent on the DNA element, it may be possible to use the assay to assess the relevance of a particular DNA-binding protein for the element's function. As a starting point, monoclonal or polyclonal antibodies directed against the candidate protein can be added to the in vitro reactions to determine whether they block the function of the control element (Fig. 9.5A). Concentrated, affinity-purified antibodies may be needed because it may be difficult to add sufficient antibody to neutralize the protein. To determine whether sufficient antibody has been added, its effect on an EMSA complex can be monitored. Control experiments should also be performed to determine whether the antibody affects transcription from an unrelated promoter that is independent of the DNA-binding protein. Additional controls are also needed, including reactions monitoring the effect of an unrelated antibody prepared by a similar method.

If antibody addition does not inhibit transcriptional activity, the candidate binding protein may not be essential for the activity of the control element. Alternatively, the antibody may not bind to an essential epitope of the protein or may not be sufficiently concentrated to neutralize all of the protein molecules within the extract.

An alternative strategy, which may be more successful, is to deplete the protein from the extract by immunoprecipitation or immunoaffinity chromatography, or by sequence-specific DNA affinity chromatography (Fig. 9.5B) (see also Chapter 8). The immunoprecipitation and immunoaffinity methods should allow the efficient depletion of a protein from an extract, unless the critical epitope of the protein is blocked by its tight association with other proteins. DNA affinity chromatography results in the depletion of all proteins that bind the element of interest, providing little insight into the identity of the protein that functionally interacts with the element. However, that insight can be provided by the subsequent addition of a recombinant or pure form of the putative regulator. If efficient transcription is restored, the results support the hypothesis that the protein is a relevant regulator of the gene.

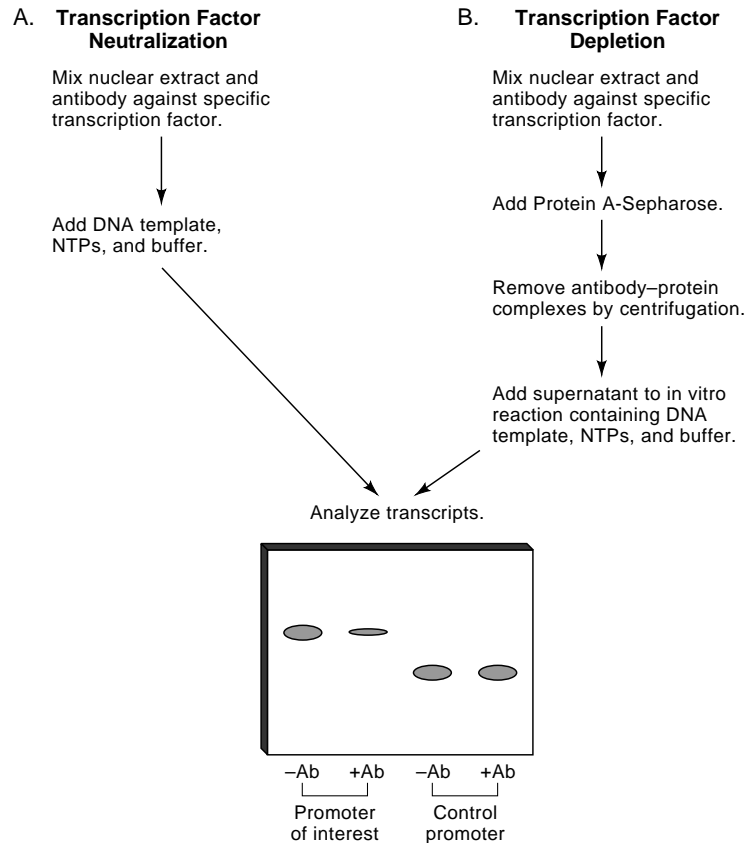


FIGURE 9.5. Assessing the relevance of a protein–DNA interaction by antibody inhibition of an in vitro transcription reaction.

In practice, depletion experiments to monitor the requirement for a protein in an in vitro transcription assay have proven to be tedious and difficult. Depletion often inhibits the in vitro transcription assay nonspecifically because of extract dilution or inactivation of a general transcription factor. The protein bound by the antibody or DNA resin also has the potential to bind or form aggregates with other proteins that are essential for in vitro transcription, resulting in the co-depletion of those proteins. These difficulties are not unusual, because the extracts must be maintained in a concentrated state during depletion to function in the subsequent in vitro reactions.

Because of the potential problems with these types of experiments, careful controls are needed, including immunodepletion controls with antibodies that should have no effect on transcription from the promoter of interest, as well as control templates that should not be affected by the depletion. Furthermore, as mentioned above, to demonstrate that the specific inhibition observed is due solely to depletion of the protein of interest, it should be possible to restore transcription by complementing the reaction with a recombinant or pure form of the protein. The results obtained using these in vitro approaches can support the functional importance of a protein for transcription of a target gene. However, the relevance of the protein would require additional support because a protein that functionally interacts with a control element in vitro is not necessarily responsible for its function in vivo.

An in vitro transcription analysis of the IFN- β enhancer provides an example of transcription factor depletion by immunoaffinity and DNA affinity chromatography (Kim and Maniatis 1997). The activity of this enhancer in vivo is thought to require ATF2/c-jun, an

IRF family member, NF- κ B, and HMGI(Y) (see Chapter 1). To study the specific protein requirements for enhancer function *in vitro*, an *in vitro* transcription assay dependent on the binding site for each factor was developed. Then, the first three factors were depleted by sequence-specific DNA affinity chromatography. Excess HMGI(Y) was depleted by immunoaffinity chromatography. To assess the function of each protein, recombinant forms of the proteins were added back to the depleted extracts, resulting in a restoration of enhancer activity. These results provided evidence that the specific proteins used are capable of activating the enhancer. However, it should be noted that recombinant IRF-1 was used for these studies, whereas more recent findings strongly suggest that other IRF family members, in particular IRF-3 and IRF-7, are actually the relevant activators of IFN- β transcription (see above and Wathelet et al. 1998). Thus, as mentioned above, the *in vitro* transcription method by itself is unreliable for demonstrating the relevance of a protein–DNA interaction.

Although this section has focused on strategies that make use of protein neutralization and depletion in crude extracts, *in vitro* transcription can also be used as an assay for the unbiased purification of a protein that activates a promoter (see Chapter 14). A classic example of this strategy led to the identification, purification, and cloning of the Sp1 protein that stimulates the SV40 early promoter (Dyan and Tjian 1983a,b; Kadonaga et al. 1987). An *in vitro* transcription assay was first developed that supported accurate transcription from the promoter. The cell extract was then fractionated to identify an essential promoter-specific transcription factor. The factor identified, Sp1, was found to bind specific DNA-sequence elements within the promoter. Sp1 was then purified, a partial peptide sequence was obtained, and its gene was cloned. The isolation of Sp1 as a protein that selectively activates the SV40 promoter provides reasonably strong support for the hypothesis that it is a relevant activator of the promoter *in vivo*.

In Vivo Protein–DNA Crosslinking

One of the strategies discussed above for confirming the importance of a protein–DNA interaction is to compare a genomic footprinting pattern to a footprinting pattern observed *in vitro* with a purified protein. This strategy can provide some evidence that a particular protein is bound to control elements within the context of the endogenous locus. One limitation of this approach, however, is that it usually cannot distinguish between the various members of a multiprotein family, which often yield indistinguishable footprinting patterns.

An alternative strategy for identifying the specific protein that associates with a control element in the context of an endogenous allele is to use *in vivo* protein–DNA crosslinking (Fig. 9.6), which was first discussed in Chapter 3. This method, originally developed as a means of determining whether RNA polymerase is paused on the transcribed leader of a gene prior to induction (see Chapter 3; Gilmour and Lis 1984, 1985, 1986), has recently been used to monitor the association of specific transcription factors with DNA (Walter et al. 1994; Boyd and Farnham 1997; Boyd et al. 1998; Wathelet et al. 1998).

In brief, growing cells are treated with ultraviolet light or formaldehyde to crosslink DNA-binding proteins to their target sites (Orlando et al. 1997; Walter and Biggin 1997). The cells are then lysed, and the DNA is cleaved into fragments by digestion with a restriction enzyme or by shearing. Protein–DNA complexes are then purified by immunoprecipitation with antibodies directed against the DNA-binding protein of interest. To determine whether the protein was crosslinked to a putative target element, the immunoprecipitate is analyzed by Southern blot or PCR for the presence of a DNA fragment encompassing the

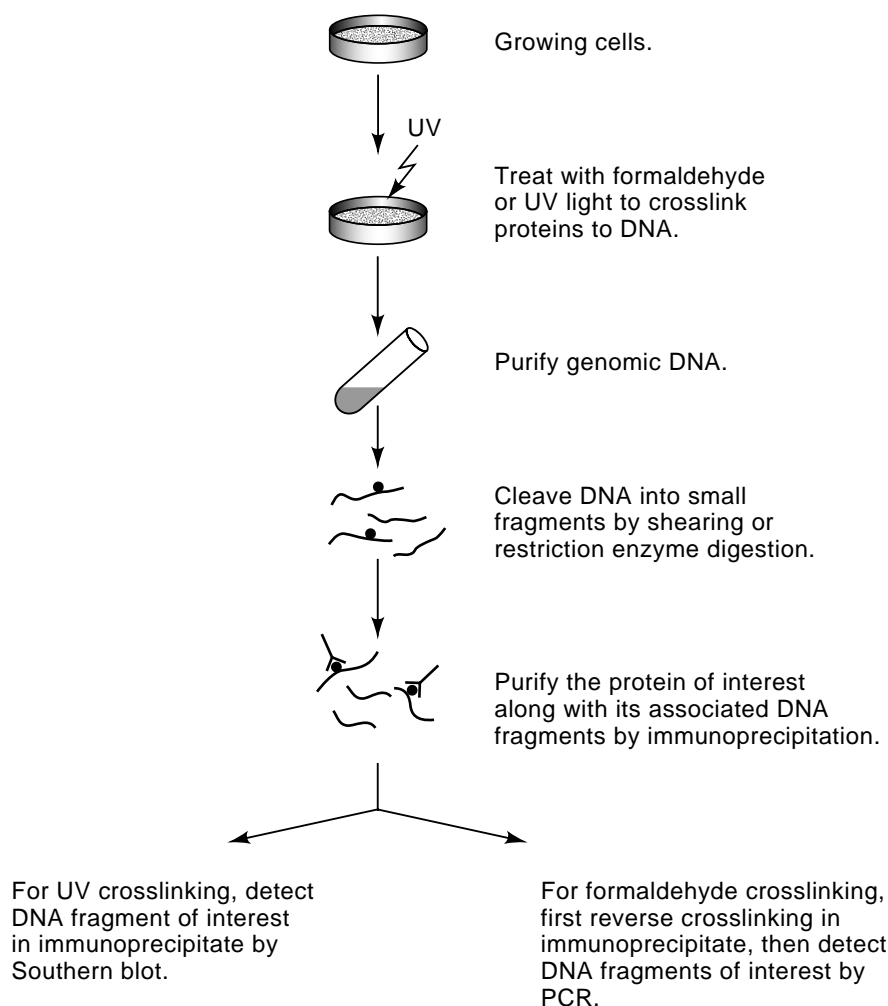


FIGURE 9.6. In vivo crosslinking of proteins to DNA-binding sites.

element. PCR is most useful as a detection method when coupled to formaldehyde crosslinking, which can be reversed following immunoprecipitation, preventing the linked protein from interfering with the PCR reaction.

The IFN- β enhancer analysis by the Maniatis laboratory provides an example of the crosslinking strategy (Fig. 9.7) (Wathelet et al. 1998). Other experiments carried out by the laboratory had implicated two IRF family members, IRF-3 and IRF-7, as relevant activators of the IFN- β enhancer, which is induced by virus infection. To strengthen the hypothesis that these two proteins are indeed relevant activators, the authors used the in vivo crosslinking strategy in mock-infected and virus-infected cells. Following formaldehyde crosslinking, extract preparation, and DNA fragmentation, antibodies against various IRF family members were used for immunoprecipitation. PCR analysis of the DNA within the immunoprecipitates revealed that the IFN- β enhancer fragment was present following immunoprecipitation with IRF-3 antibodies and IRF-7 antibodies, but not following immunoprecipitation with IRF-1 antibodies (Fig. 9.7). Importantly, the IFN- β enhancer fragment was only present in the immunoprecipitates from virus-infected cells and not

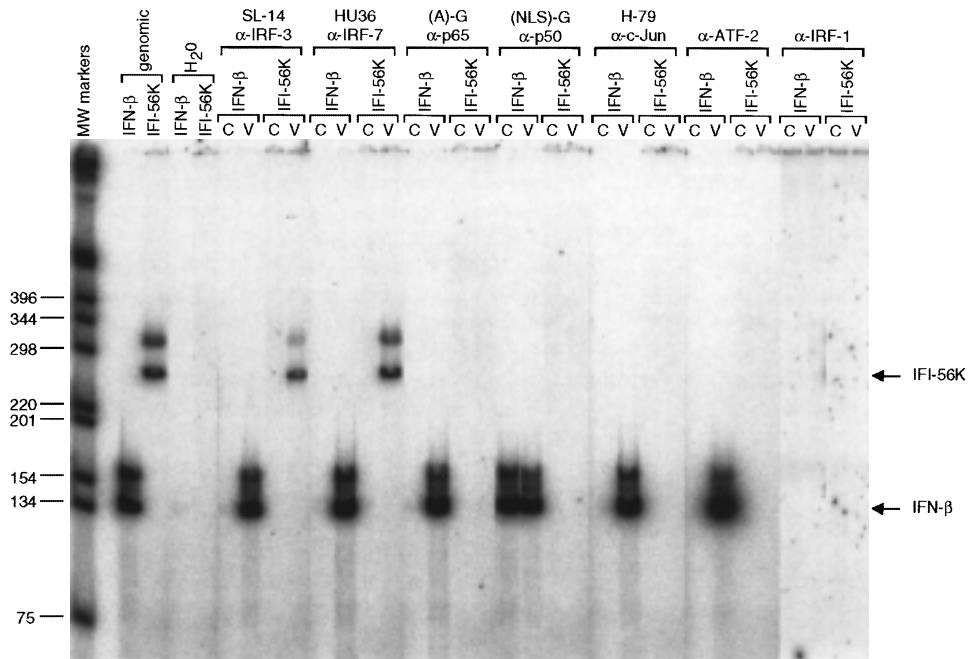


FIGURE 9.7. Association of specific transcriptional activator proteins with virus-inducible promoters *in vivo*. HeLa cells were mock (C)- or SV (V)-infected for 6 hours, and treated with formaldehyde to crosslink proteins bound to DNA. Crosslinks were purified and immunoprecipitated with the indicated antibodies as described previously (Orlando et al. 1997). After reversal of the crosslinking, the DNA was amplified using primers for the IFI-56K (–237/–219 and +12/+31) and IFN- β (–129/–117 and –1/–19) promoters (PCRs, 20 cycles of 1 min at 94°C, 55°C, and 72°C each). Molecular-weight marker and control amplification using genomic DNA and water are shown on the left. (Reprinted, with permission, from Wathelet et al. 1998 [Copyright 1998 Cell Press].)

from mock-infected cells. Immunoprecipitation with antibodies against other DNA-binding proteins (p65, p50, c-Jun, and ATF-2) confirmed that they also interact with the IFN- β promoter, but not with a control promoter (IFI-56K, Fig. 9.7). These results provide strong evidence that IRF-3 and IRF-7 are indeed relevant activators of IFN- β enhancer activity.

The principal strength of the *in vivo* crosslinking strategy is that it is the only method currently available for directly “visualizing” an *in vivo* interaction between a specific protein and control element. If compelling data are obtained, the method can provide strong and direct evidence that a site is occupied by a specific protein in growing cells.

A limitation of the approach is that it is difficult to obtain compelling data. In fact, the procedure had been attempted in mammalian cells for several years, but manuscripts making use of it have only recently begun to appear in the literature (see, e.g., Boyd and Farnham 1997; Boyd et al. 1998; Wathelet et al. 1998). The principal controls needed for convincing results are (1) experiments performed with multiple unrelated antibodies, which should not precipitate the DNA fragment of interest, and (2) experiments with PCR primer sets or Southern blot probes directed against several other chromosomal regions that should not precipitate with the specific antibodies. An additional limitation of the crosslinking strategy is that association of a protein with a specific DNA fragment does not provide conclusive evidence that the association is functionally relevant. In fact, Walter et al. (1994) have reported that a subset of *Drosophila* homeodomain proteins crosslink to a large number of genomic sites, whereas other homeodomain proteins crosslink to a small number of genes that are thought to be direct targets. The reason for the relatively ubiqui-

tous crosslinking of some proteins remains unknown. However, these results emphasize the need for careful controls and for the cautious interpretation of crosslinking data.

Altered Specificity Experiments

The final strategy that is discussed for testing the relevance of a protein–DNA interaction is an altered specificity strategy (Fig. 9.8). Although an ideal altered-specificity experiment is difficult to design and perform, it has the potential to provide more compelling evidence that a protein–DNA interaction is relevant than any of the other strategies described above. To perform this type of experiment, the DNA-binding domain of the protein of interest is first mutated so that it recognizes a different DNA sequence. The new sequence recognized by the altered protein is then inserted into the control region of interest in place of the sequence element recognized by the wild-type protein. It is important for the new recognition sequence to be fairly unique, so that it is not recognized with a significant affinity by other DNA-binding proteins within the cell. The altered-specificity DNA-binding protein is then expressed in cells containing an endogenous gene or reporter gene regulated by the altered control region, and its capacity to regulate transcription is monitored.

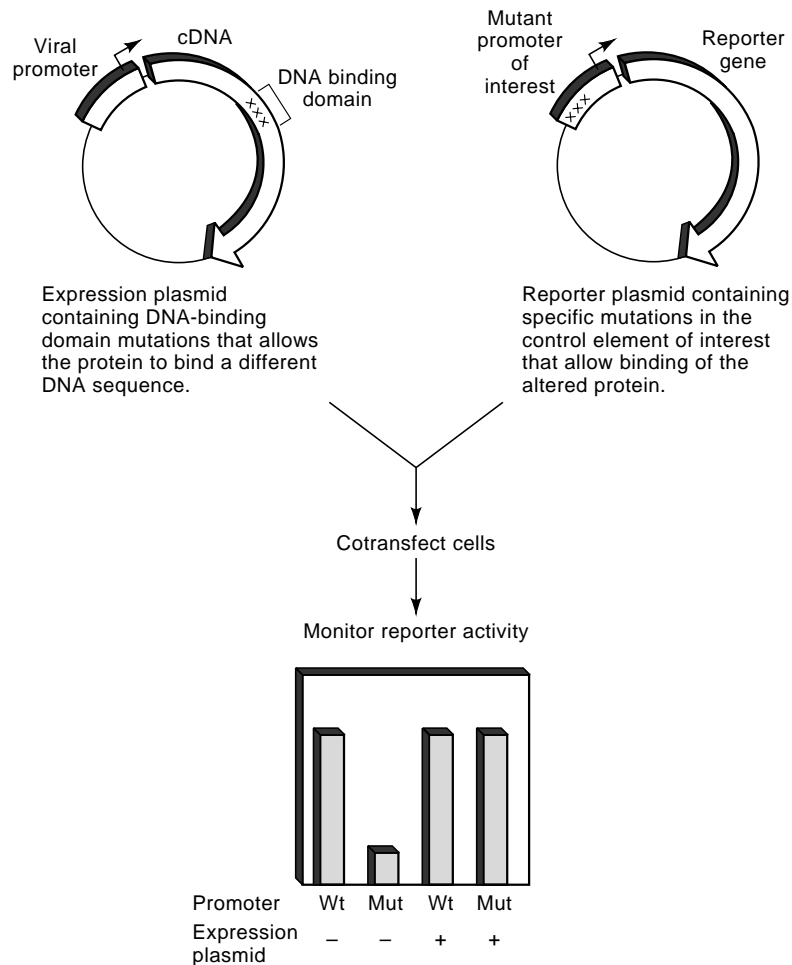


FIGURE 9.8. Altered specificity strategy.

A particularly elegant example of an altered-specificity experiment, which reveals its strengths and weaknesses, was reported by Shah et al. (1997). (Another important example was recently reported by Gillemans et al. 1998.) The goal of this study was to determine which of two POU-domain proteins, OCT-1 or OCT-2, is a functional activator of the Ig heavy- and light-chain gene promoters. Both proteins can bind with similar affinity to the octamer elements within the Ig promoters, both proteins are expressed in B lymphocytes, and gene disruption experiments have been uninformative. Therefore, it has been difficult to identify the relevant activator of the Ig promoters.

To create an altered-specificity OCT protein, Shah et al. (1997) examined the known crystal structure of the OCT-DNA complex. They focused their attention on a particular amino acid sequence within the POU domain that contacted one of the nucleotides within the octamer DNA sequence. They first mutated that nucleotide, which disrupted binding by the wild-type OCT proteins. Then, they set out to isolate an altered OCT POU domain capable of binding the altered DNA sequence with high affinity. To achieve this goal, they generated a phage expression library for the POU domain, in which each phage expressed a POU domain with a randomly generated amino acid sequence in the region that was predicted to be in close proximity to the altered base pair. To isolate phage that express a mutant POU domain capable of binding the altered DNA sequence, the expression library was probed with a radiolabeled oligonucleotide containing the altered sequence. The amino acids within the selected POU domain, which allowed high-affinity binding, were then determined and introduced into mammalian expression plasmids for both OCT-1 and OCT-2.

The altered OCT-1 and OCT-2 proteins were tested in B cells for their ability to *trans*-activate reporter plasmids under the control of Ig promoters containing the altered octamer DNA sequence. By using the altered octamer sequence, endogenous OCT-1 and OCT-2 proteins within the B cells were rendered nonfunctional on the Ig promoters; only the altered-specificity OCT-1 or OCT-2 introduced into the cells could bind. Because the experiments could be performed in B cells, it was anticipated that the altered proteins would be capable of functionally interacting with the other proteins needed for B-cell-specific Ig promoter and enhancer activity.

The results revealed that the altered OCT-1 and OCT-2 proteins were equally capable of stimulating Ig promoter activity. Surprisingly, however, when the reporter plasmids included an Ig enhancer in addition to the Ig promoter, OCT-1 was found to be a much more potent activator. These results are consistent with a hypothesis in which OCT-1 is the functional activator of endogenous Ig genes, because only OCT-1 can carry out the protein-protein interactions needed for promoter- and enhancer-dependent transcription.

The altered specificity strategy can provide compelling evidence that a DNA-binding protein acts at a particular target site. In effect, this strategy can provide much of the same information as would be provided by a gene disruption experiment (see above), but the key limitations of a gene disruption experiment are eliminated. One important limitation of gene disruption is that loss of the DNA-binding protein might result in a loss of cell viability. A second limitation is that it is extremely difficult to distinguish between direct and indirect effects of the binding protein on the candidate target gene. A third limitation is that the DNA-binding protein may be redundant with a related protein, such that its disruption has no effect on target gene transcription. None of these limitations is relevant in the altered-specificity experiment: The cells remain viable because the endogenous DNA-binding protein is still expressed; indirect effects are less likely because the altered recognition site for the altered-specificity protein has been introduced only into the control region

of interest, and redundancy is not observed because the altered DNA-binding domain is present in only one protein.

Despite these considerable advantages, the altered-specificity approach has three limitations. One limitation is that it typically involves overexpression of the altered DNA-binding protein by either transient or stable transfection. As described above, results obtained with overexpressed proteins are inconclusive and difficult to interpret. To enhance the degree to which altered-specificity results can be interpreted, the altered protein should be expressed at a concentration similar to that of the endogenous, wild-type protein. To achieve this goal, different stable cell lines expressing the altered protein can be examined, or a variety of expression vectors containing different promoters or enhancers can be tested.

The second limitation arises when the altered control region is analyzed in the context of a transiently transfected reporter plasmid. The high copy number of the reporter plasmid and its removal from its natural chromosomal environment could influence its ability to respond to the altered-specificity protein. In an ideal experiment, the substitution mutation creating the altered control element would be introduced into the endogenous gene by homologous recombination. This rigorous approach would provide the strongest evidence that a particular DNA-binding protein acts on a target gene. Of course, this experiment would require much more effort than the basic transfection experiment. The use of a stably transfected reporter plasmid would be preferred to the use of a transiently transfected reporter, but the results would remain inconclusive.

The third limitation, which cannot be overcome, is that the transcriptional activation function of the wild-type DNA-binding protein might depend on the specific amino acids that are altered. These amino acids could be directly involved in transcriptional activation, or the protein–DNA interaction could lead to a conformational change in another surface of the protein involved in transcriptional activation. In most altered-specificity DNA-binding proteins, the amino acids responsible for DNA binding are unimportant for transcriptional activation, but examples of a conformational link between DNA binding and activation have been described (Lefstin and Yamamoto 1998).

Nevertheless, despite this limitation and the others mentioned above, the altered-specificity strategy can provide valuable information regarding the relevance of a protein–DNA interaction and holds an important position among the arsenal of strategies described in this chapter.

To summarize, we have described twelve general strategies that can be used to test the hypothesis that a DNA-binding protein is a relevant regulator of a gene by binding a defined control element. Some of these strategies may not be feasible for analysis of some DNA-binding proteins, and other strategies that were not described can be envisioned (e.g., *in vivo* antibody microinjection experiments). As stated in the introduction, no single strategy can conclusively establish the functional relevance of a protein–DNA interaction. However, a hypothesis can be greatly strengthened by subjecting it to as many rigorous tests as possible.

REFERENCES

- Benezra R., Davis R.L., Lockshon D., Turner D.L., and Weintraub H. 1990. The protein Id: A negative regulator of helix-loop-helix binding proteins. *Cell* **61**: 49–59.
- Boyd K.E. and Farnham P.J. 1997. Myc versus USF: Discrimination at the cad gene is determined by core promoter elements. *Mol. Cell. Biol.* **17**: 2529–2537.
- Boyd K.E., Wells J., Gutman J., Bartley S.M., and Farnham P.J. 1998. c-Myc target gene specificity is determined by a post-DNA binding mechanism. *Proc. Natl. Acad. Sci.* **95**: 13887–13892.

- Branch A.D. 1998. A good antisense molecule is hard to find. *Trends Biochem. Sci.* **23**: 45–50.
- Capecchi M.R. 1989. Altering the genome by homologous recombination. *Science* **244**: 1288–1292.
- Clevers H.C. and Grosschedl R. 1996. Transcriptional control of lymphoid development: Lessons from gene targeting. *Immunol. Today* **17**: 336–343.
- Descombes P. and Schibler U. 1991. A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. *Cell* **67**: 569–579.
- Dynan W.S. and Tjian R. 1983a. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* **32**: 669–680.
- . 1983b. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35**: 79–87.
- Eisenbeis C.F., Singh H., and Storb U. 1995. Pip, a novel IRF family member, is a lymphoid-specific, PU.1-dependent transcriptional activator. *Genes Dev.* **9**: 1377–1387.
- Ernst P., Hahm K., and Smale S.T. 1993. Both LyF-1 and an Ets protein interact with a critical promoter element in the murine terminal transferase gene. *Mol. Cell. Biol.* **13**: 2982–2992.
- Ernst P., Hahm K., Trinh L., Davis J.N., Roussel M.F., Turck C.W., and Smale S.T. 1996. A potential role for Elf-1 in terminal transferase gene regulation. *Mol. Cell. Biol.* **16**: 6121–6131.
- Fiering S., Northrop J.P., Nolan G.P., Mattila P.S., Crabtree G.R., and Herzenberg L.A. 1990. Single cell assay of a transcription factor reveals a threshold in transcription activated by signals emanating from the T-cell antigen receptor. *Genes Dev.* **4**: 1823–1834.
- Georgopoulos K., Moore D.D., and Derfler B. 1992. Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. *Science* **258**: 808–812.
- Georgopoulos K., Bigby M., Wang J.H., Molnar A., Wu P., Winandy S., and Sharpe A. 1994. The Ikaros gene is required for the development of all lymphoid lineages. *Cell* **79**: 143–156.
- Gillemans N., Tewari R., Lindeboom F., Rottier R., de Wit T., Wijgerde M., Grosveld F., and Philipsen S. 1998. Altered DNA-binding specificity mutants of EKLF and Sp1 show that EKLF is an activator of the β -globin locus control region in vivo. *Genes Dev.* **12**: 2863–2873.
- Gilmour D.S. and Lis J.T. 1984. Detecting protein-DNA interactions in vivo: Distribution of RNA polymerase on specific bacterial genes. *Proc. Natl. Acad. Sci.* **81**: 4275–4279.
- . 1985. In vivo interactions of RNA polymerase II with genes of *Drosophila melanogaster*. *Mol. Cell. Biol.* **5**: 2009–2018.
- . 1986. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Mol. Cell. Biol.* **6**: 3984–3989.
- Hahm K., Ernst P., Lo K., Kim G.S., Turck C., and Smale S.T. 1994. The lymphoid transcription factor LyF-1 is encoded by specific, alternatively-spliced mRNAs derived from the *Ikaros* gene. *Mol. Cell. Biol.* **14**: 7111–7123.
- Jonsen M.D., Petersen J.M., Xu Q.P., and Graves B.J. 1996. Characterization of the cooperative function of inhibitory sequences in Ets-1. *Mol. Cell. Biol.* **16**: 2065–2073.
- Kadonaga J.T., Carner K.R., Masiarz F.R., and Tjian R. 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**: 1079–1090.
- Kim T.K. and Maniatis M. 1997. The mechanism of transcriptional synergy of an in vitro assembled interferon- β enhanceosome. *Mol. Cell* **1**: 119–129.
- Langlands, K., Yin X., Anand G., and Prochownik E.V. 1997. Differential interactions of Id proteins with basic helix-loop-helix transcription factors. *J. Biol. Chem.* **272**: 19785–19793.
- Lauring J. and Schlissel M.S. 1999. Distinct factors regulate the murine RAG-2 promoter in B and T cell lines. *Mol. Cell. Biol.* **19**: 2601–2612.
- Lefstin J.A. and Yamamoto K.R. 1998. Allosteric effects of DNA on transcriptional regulators. *Nature* **392**: 885–888.
- Lehman A.M., Ellwood K.B., Middleton B.E., and Carey M. 1998. Compensatory energetic relationships between upstream activators and the RNA polymerase II general transcription machinery. *J. Biol. Chem.* **273**: 932–939.
- Lo K., Landau N.R., and Smale S.T. 1991. LyF-1, a transcriptional regulator that interacts with a novel class of promoters for lymphocyte-specific genes. *Mol. Cell. Biol.* **11**: 5229–5243.

- Lodish H., Baltimore D., Berk A., Zipursky S.L., and Darnell J. 1996. *Molecular cell biology*, 3rd edition. W.H. Freeman, New York.
- Mauxion F., Jamieson C., Yoshida M., Arai K., and Sen R. 1991. Comparison of constitutive and inducible transcriptional enhancement mediated by κ B-related sequences: Modulation of activity in B cells by human T-cell leukemia virus type I tax gene. *Proc. Natl. Acad. Sci.* **88**: 2141–2145.
- Murphy T.L., Cleveland M.G., Kulesza P., Magram J., and Murphy K.M. 1995. Regulation of interleukin 12 p40 expression through an NF- κ B half-site. *Mol. Cell. Biol.* **15**: 5258–5267.
- Okamura R.M., Sigvardsson M., Galceran J., Verbeek S., Clevers H., and Grosschedl R. 1998. Redundant regulation of T cell differentiation and TCR α gene expression by the transcription factors LEF-1 and TCF-1. *Immunity* **8**: 11–20.
- Orlando V., Strutt H., and Paro R. 1997. Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods* **11**: 205–214.
- Plevy S.E., Gemberling J., Hsu S., Dorner A.J., and Smale S.T. 1997. Multiple control elements mediate activation of the murine and human IL-12 p40 promoters: Evidence of functional synergy between C/EBP and Rel proteins. *Mol. Cell. Biol.* **17**: 4572–4588.
- Pongubala J.M., Nagulapalli S., Klemsz M.J., McKercher S.R., Maki R.A., and Atchison M.L. 1992. PU.1 recruits a second nuclear factor to a site important for immunoglobulin κ 3' enhancer activity. *Mol. Cell. Biol.* **12**: 368–378.
- Shah P.C., Bertolino E., and Singh H. 1997. Using altered specificity Oct-1 and Oct-2 mutants to analyze the regulation of immunoglobulin gene transcription. *EMBO J.* **16**: 7105–7117.
- Stein C.A. 1998. How to design an antisense oligodeoxynucleotide experiment: A consensus approach. *Antisense Nucleic Acid Drug Dev.* **8**: 129–132.
- Thanos D. and Maniatis T. 1995. NF- κ B: A lesson in family values. *Cell* **80**: 529–532.
- van Genderen C., Okamura R.M., Farinas I., Quo R.G., Parslow T.G., Bruhn L., and Grosschedl R. 1994. Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes Dev.* **8**: 2691–2703.
- Verbeek S., Izon D., Hofhuis F., Robanus-Maandag E., te Riele H., van de Wetering M., Oosterwegel M., Wilson A., MacDonald H.R., and Clevers H. 1995. An HMG-box-containing T-cell factor required for thymocyte differentiation. *Nature* **374**: 70–74.
- Walter J. and Biggin M.D. 1997. Measurement of in vivo DNA binding by sequence-specific transcription factors using UV cross-linking. *Methods* **11**: 215–224.
- Walter J., Dever C.A., and Biggin M.D. 1994. Two homeo domain proteins bind with similar specificity to a wide range of DNA sites in *Drosophila* embryos. *Genes Dev.* **8**: 1678–1692.
- Wathelet M.G., Lin C.H., Parekh B.S., Ronco L.V., Howley P.M., and Maniatis T. 1998. Virus infection induces the assembly of coordinately activated transcription factors on the IFN- β enhancer in vivo. *Mol. Cell* **1**: 507–518.

