



Characterizing transcription factor binding sites using formaldehyde crosslinking and immunoprecipitation

Julie Wells and Peggy J. Farnham*

McArdle Laboratory for Cancer Research, University of Wisconsin Medical School, 1400 University Avenue, Madison, Wisconsin 53706, USA

Accepted 15 January 2002

Abstract

In their article, A.S. Weinmann and P.J. Farnham (2002, *Methods* 26) described new techniques for isolating *in vivo* binding sites for any DNA-binding protein. In this article, we describe complementary methods for detailed *in vivo* characterizations of such identified protein–DNA interactions. First, we describe how formaldehyde crosslinking and chromatin immunoprecipitation (ChIP), in conjunction with transient transfections or the use of cell lines containing stably integrated constructs or episomes, can be employed to identify which specific nucleotides of a region of DNA are required for recruitment of a particular transcription factor. In contrast to *in vivo* footprinting, this method not only specifies which nucleotides are bound, but also identifies the protein(s) involved in binding. Next, we discuss the use of the ChIP assay to study how binding of a transcription factor is altered by passage through the cell cycle, by overexpression or deletion of another factor, or during tumorigenesis. Finally, a look toward the future suggests that the ChIP assay may be combined with Western blot analysis or mass spectrometry to identify additional proteins that interact with a transcription factor of interest. © 2002 Elsevier Science (USA). All rights reserved.

Keywords: Crosslinking; Immunoprecipitation; Transcription factor; Cell cycle regulation; Tumorigenesis; Tissues; Synchronization; Protein–protein interactions

1. Introduction

The search for target genes of an individual transcription factor has always been an area of interest but the search has become more feasible with the recent development of techniques, such as SAGE and DNA microarrays, that allow large numbers of possible target genes to be screened at once. However, these techniques identify both direct targets (i.e., those genes whose expression is regulated by the transcription factor of interest), as well as indirect targets (i.e., those genes whose expression is altered as consequence of a change in the cellular environment). Because these techniques usually require overexpression or deletion of the transcription factor of interest, they do not necessarily identify genes that are targets under unperturbed, physiological conditions. Furthermore, it would be difficult to use these techniques to find targets under altered conditions such

as during different stages of development or during tumorigenesis.

In their article, Weinmann and Farnham [14] describe two techniques for identifying *in vivo* binding sites of a transcription factor that binds DNA either directly (via its own DNA binding domain) or indirectly (as a coactivator or corepressor). Both of these techniques are based on formaldehyde crosslinking and chromatin immunoprecipitation (ChIP). In brief, ChIP involves using formaldehyde to chemically crosslink bound proteins to underlying DNA sequences or to crosslink proteins of multiprotein complexes together. The crosslinked chromatin is sonicated to achieve the desired size and then DNA fragments and protein complexes of interest are immunoprecipitated with an antibody against one of the crosslinked proteins. As described by Weinmann and Farnham [14], the immunoprecipitated fragments can be cloned into a suitable vector and sequenced or labeled with a fluorescent dye and hybridized to a promoter-enriched DNA array. Following this initial identification and subsequent confirmation of *in vivo* binding sites, the next step then

* Corresponding author. Fax: (608) 262-2824.

E-mail address: farnham@oncology.wisc.edu (P.J. Farnham).

becomes a further characterization of the positive binding sites. In this article, we describe how the ChIP assay, alone or in combination with other molecular biology techniques, can be used to identify precisely which nucleotides within a region of DNA are required for *in vivo* binding of the transcription factor of interest. Additionally, we describe the use of the ChIP assay to study how binding of a transcription factor to a promoter can be influenced by cell cycle phase, by the relative concentrations of other proteins, or during tumor development.

2. Description of method

2.1. Which nucleotides are required for *in vivo* binding?

The two techniques described by Weinmann and Farnham identify approximately 1- to 3-kb regions of DNA that contain at least one binding site for the factor of interest. Because it is difficult to shear chromatin into fragments of less than several hundred base pairs, standard ChIP is not a suitable assay for high-resolution characterization of binding sites. At present, *in vitro* gel mobility shift analysis using cellular extract is the standard assay for determining which specific nucleotides within a region of DNA constitute a binding site for a particular factor. Although gel mobility shift analysis is a fast and easy method for identifying which nucleotides are required for binding under *in vitro* conditions, *in vivo* binding conditions are often difficult to recreate *in vitro* (see Fig. 1). For example, binding of proteins that require DNA secondary structure, such as looping, to bring two distal binding sites into close proximity often cannot be studied using this method. Additionally, sites that require multiprotein complex formation to stabilize protein–DNA interactions are difficult to study *in vitro* due to dissolution of the complex during gel electrophoresis. Finally, if a transcription factor is recruited to DNA via interaction with another, perhaps unknown, DNA-binding protein *in vitro* assay conditions may not

be optimized for binding of the factor that actually contacts DNA.

An example of a binding site that is refractory to *in vitro* analysis comes from the ChET8 promoter. ChET8 was originally identified in a screen for E2F1 target genes and was confirmed as containing a strong *in vivo* binding site for multiple E2Fs using the standard ChIP assay [1]. In subsequent gel shift analysis, however, ChET8 did not efficiently compete for binding of E2F complexes to a probe containing a consensus E2F binding site. Inspection of the entire sequence that was isolated by ChIP cloning indicates that ChET8 does not contain a consensus or near-consensus E2F binding site. These results suggest either that E2F is binding to a nonconsensus E2F binding site within the ChET8 fragment or that E2F is being recruited to ChET8 through interaction with another DNA-binding protein. Regardless of which mechanism allows E2F to occupy the ChET8 promoter *in vivo*, *in vitro* analysis has not been useful in defining the E2F binding site within this promoter. An alternative approach to define the E2F binding site of the ChET8 fragment is to use *in vivo* footprinting. *In vivo* footprinting offers a distinct advantage over gel mobility shift analysis in that protein binding can be studied in living cells. However, *in vivo* footprinting is capable only of identifying regions of the DNA that are bound by protein; this technique cannot be used to characterize the protein complex that is responsible for the observed footprint. In the case of ChET8, it is possible that a consensus element for another DNA-binding protein may be the site of E2F recruitment. However, using *in vivo* footprinting one could never be sure whether the identified site recruited the other factor alone or the other factor plus E2F.

In an effort to address the shortcomings of both *in vitro* gel mobility shift assays and *in vivo* footprinting, we and others have modified the standard ChIP assay to monitor transcription factor binding to transfected DNA fragments. For example, we have created cell lines containing stably integrated promoter–reporter constructs. By comparing transcription factor binding to

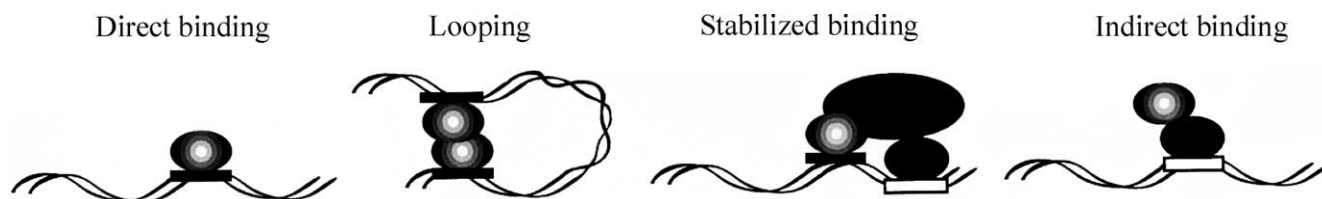


Fig. 1. Four possible mechanisms for protein binding. A transcription factor (shaded circle) can bind to target sites through one of at least four distinct mechanisms. First, a transcription factor can be in direct contact with either a consensus or nonconsensus binding site (filled rectangle). Alternatively, two or more transcription factor molecules can bind to separated sites that are brought into close proximity by looping of the intervening DNA via protein–protein interactions. Binding of a transcription factor to a low-affinity site can be stabilized by contact with another protein (large dark oval) which is recruited to the DNA via a second site-specific transcription factor (small dark oval) bound to a different site (open rectangle). Finally, proteins can be brought to sites “piggyback” fashion through interaction with other DNA-binding proteins (small dark oval) bound to different binding sites.

Table 1
Advantages and disadvantages of three methods for determining which nucleotides are required for in vivo binding^a

Method	Advantages	Disadvantages
Transient transfection	Quick Easy	Difficult to control copy number Insert not surrounded by chromatin Limited insert size
Stable integration	Insert located within chromatin	Time consuming Effects of different sites of integration Limited insert size
Episomes	Possible to work with very large inserts or entire genes	Time consuming Limited choice of cell lines

^aTransient transfection, stable integration, and introduction of episomes can all be used to generate cells harboring constructs to determine which nucleotides of a fragment of DNA are required for in vivo binding of a protein of interest. The decision as to which method to choose depends on factors such as speed, ease, size of the fragment to be analyzed, and importance of surrounding chromatin.

constructs containing wild-type sequence versus constructs containing mutated or deleted sequences, we can precisely identify transcription factor binding sites. Furthermore, incorporation of a reporter gene, such as luciferase, into the construct allows binding of the transcription factor to be correlated with promoter activity. An added benefit of using ChIP with transiently transfected cells to identify your binding site is that this type of analysis is amenable to a larger range of fragment sizes than is gel mobility shift analysis. In place of cells containing stably integrated constructs, transiently transfected cells or cells containing episomes can also be used. All three of these methods (stably integrated plasmids, transiently transfected plasmids, and episomes) have differing advantages and shortcomings which are listed in Table 1.

All three methods of binding site characterization rely on transfecting into cells constructs containing overlapping fragments of a region of DNA, such as a 1- to 3-kb region identified in one of the screens described in the article by Weinmann and Farnham [14], and then performing ChIP analysis to determine if the transcription factor of interest still binds to the introduced fragment (Fig. 2). Once the region containing the binding site has been sufficiently narrowed, the process can be repeated with fragments containing point mutations within or deletions of the suspected binding site to identify exactly which nucleotides are required for binding. Computer programs that identify potential protein binding sites can aid in the initial process of narrowing possible regions to test. However, reliance on the identification of consensus binding sites is not always appropriate. In our laboratory, analysis of hundreds of potential target genes has shown that not all close matches to consensus E2F binding sites are bound in vivo; in contrast, some nonconsensus binding sites are occupied [1,2]. Similarly, the *cdc2* promoter contains at least two potential E2F binding sites: one site is a perfect match to the consensus binding site and the other is 1-bp mismatch. In vivo footprinting analysis showed that the perfect consensus

site was not occupied while there was a footprint over the other site and binding to that site was cell cycle dependent [3].

An example of the use of episomes to study binding of E2F and Ets to the *c-myc* promoter is shown in Fig. 3. For these experiments, Raji cells were stably transfected with one of two different episomes. “Wild-type” cells contain approximately 25–30 copies per cell of an episome bearing an 8-kb segment of the *c-myc* gene. The “E2F mutant” cell line is identical to the “wild-type” cell line except that 6 bp in the E2F binding site have been replaced with the 6 bp *EcoRI* recognition site (Fig. 3A). Primer sequences were chosen to hybridize to the portion of the *c-myc* promoter containing either the wild-type or mutated E2F binding site. Control polymerase chain reaction (PCR) analyses of the episomes using the construct-specific primers in combination with a common downstream primer indicated that the primer pairs amplify only the correct construct (Fig. 3B). Subsequent ChIP analysis revealed that E2F1, E2F2, and E2F4, but not Ets, bind to the wild-type *c-myc* promoter (Fig. 3C). E2F mutant cells, however, displayed decreased binding of E2F1, E2F2, and E2F4 and increased binding of Ets. These results are in agreement with previous studies that showed that both E2F and Ets can bind to the *c-myc* promoter, but that binding of one excludes binding of the other [4].

2.2. What protein domains are required for binding?

In addition to identifying the exact nucleotides that are required for binding of a particular transcription factor, it is also possible to use the ChIP assay to identify which domains of a protein are required for binding in vivo. This can be accomplished by the creation of cell lines that express the protein of interest with various protein domains either mutated or absent. The protein construct should contain an epitope tag, such as influenza hemagglutinin HA. An antibody against the epitope tag could then be used for subsequent ChIP

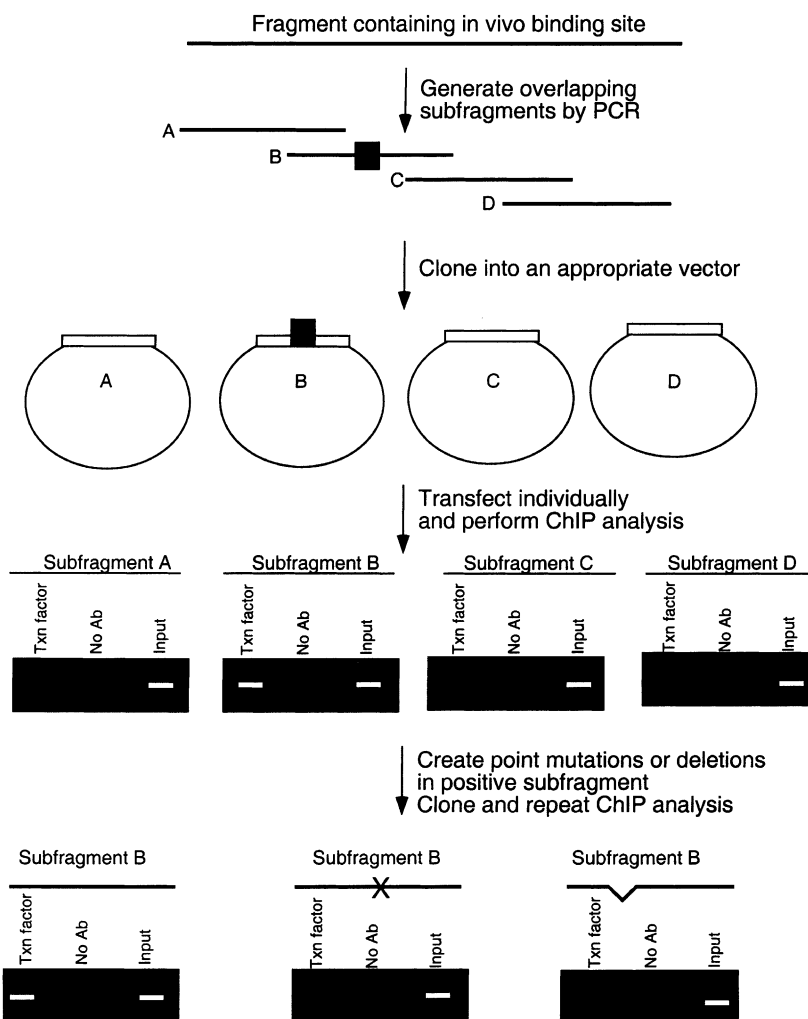


Fig. 2. Schematic representation of how to use the ChIP assay to identify precisely which nucleotides are required for in vivo binding. A region of DNA containing a transcription factor binding site (filled rectangle) at an unknown location is divided into smaller, overlapping subfragments. The subfragments are then cloned into an appropriate vector and transfected individually into cells. ChIP analysis reveals which subfragment (in this example, subfragment B) contains a binding site, as indicated by a PCR product in the sample that was immunoprecipitated with an antibody against the transcription factor of interest. Point mutations or deletions are created in the suspected binding site and ChIP analysis is repeated. If mutation or deletion of a nucleotide results in loss of PCR signal following ChIP analysis, then that nucleotide is required for in vivo binding.

analysis. Because the exogenous protein will be competing with the endogenous protein for binding, it is advisable to remove or decrease the amount of wild-type protein present if possible. This can be accomplished by using cells from an animal that is nullizygous for the protein under study, by using cells that normally contain small amounts of the wild-type protein, or by inactivating the endogenous protein. For example, in an effort to determine the role of the 52 repeats of the C-terminal domain (CTD) of RNA polymerase II, Chapman and Eick have created cell lines that express HA-tagged versions of RNA polymerase II in which various CTD repeats have been deleted. Their RNA polymerase constructs are α -amanitin resistant, allowing them to inactivate endogenous RNA polymerase II by adding α -amanitin to the tissue culture medium (R. Chapman,

and D. Eick, personal communication). Such studies will allow the involvement of individual heptapeptide repeats in recruitment of RNAP II to different promoters to be assayed.

Similar approaches that allow the requirement of specific protein domains to be monitored will also aid in distinguishing among the different mechanisms of recruitment of a transcription factor to an in vivo binding site (Fig. 1). For example, direct binding, to either a consensus or nonconsensus site, would require the DNA binding domain (DBD) of the protein in question. Alternatively, indirect binding may not require the DBD but would require an additional protein–protein interaction domain, whereas looping would require both the DBD and a protein–protein interaction domain.

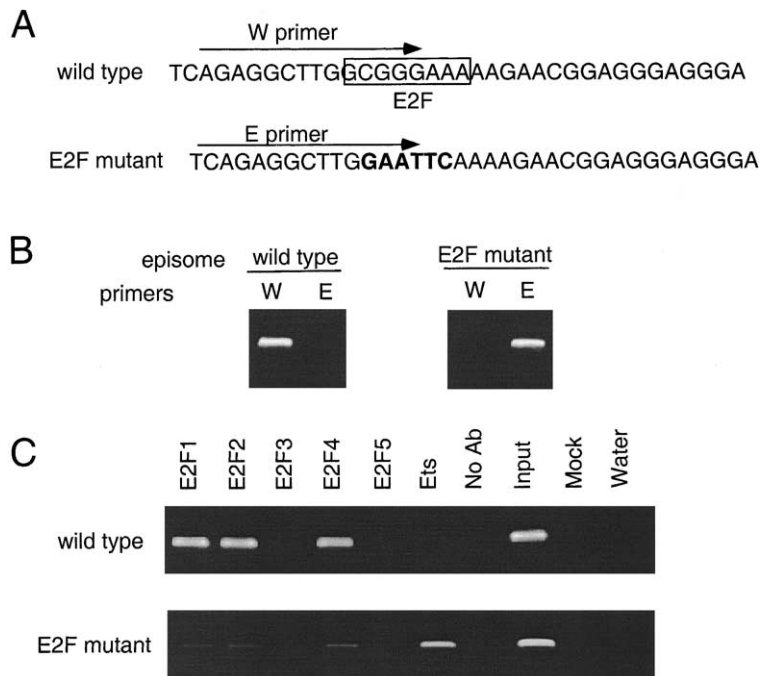


Fig. 3. Using episomes to study E2F binding to the *c-myc* promoter. (A) Schematic showing a short region of the P1 promoter of the human *c-myc* gene. The location of the E2F binding site (open rectangle) in the wild-type *c-myc* promoter, as well as the location of the six-nucleotide substitution (GAATTC) in the E2F binding site mutant episome, is indicated. Also indicated (filled arrows) are the locations of the primers used to analyze binding to this region of the *c-myc* promoter in cell lines containing either the wild-type or E2F mutant episomes. (B) Control PCR analyses of the two episome-containing cell lines using a common downstream primer plus one of the two primers indicated in (A). As expected, the wild-type primer (W) detects DNA only in the cells containing the wild-type episome and the E2F mutant primer (E) detects signal only in the cells containing the mutant episome. (C) ChIPs analysis of cell lines containing either the wild-type (top) or E2F mutant (bottom) episome using antibodies against E2F1–5 and Ets. As negative controls, the No Ab sample was an immunoprecipitation that did not contain antibody, the mock sample was an immunoprecipitation that contained buffer in place of chromatin, and the water sample was a PCR that did not contain DNA. The input sample contains 0.02% of the total starting chromatin. The appropriate primers, described in (B), were used to amplify the immunoprecipitates using 27 PCR cycles. Reprinted, with permission, from Albert et al. [12].

2.3. How can *in vivo* binding of a transcription factor be altered?

Once you have precisely identified the *in vivo* binding site of the transcription factor of interest and defined the protein domains required for the interaction, a logical question to ask next is what regulates binding of your factor to the site(s) you have identified. For example: (1) Is binding of your factor cell cycle regulated? (2) Is binding tissue specific? (3) Is binding affected by loss or overexpression of other family members? (4) Does binding vary during neoplastic transformation?

Because the ChIP assay works very well with cells synchronized by a variety of methods [i.e., serum starvation, drugs, or fluorescence-activated cell sorting (FACS)] it is a useful technique for studying cell cycle-regulated binding patterns [5]. Cells can be synchronized, crosslinked, and then either stored at 4 °C (when stored for less than 8 h) or flash-frozen in liquid nitrogen and stored at –80 °C (for up to several weeks) until all the samples have been collected. Once the cells have been formaldehyde-fixed, they do not continue to progress through the cell cycle. All of the samples can then

be processed together, eliminating experiment-to-experiment variability. Since fixation with formaldehyde is relatively quick (10 min), time points can be collected in rapid succession to study short-lived binding events. An example of the use of serum starvation and stimulation to synchronize NIH 3T3 cells for analysis of E2F binding to the Rb promoter is shown in Fig. 4A.

In addition to analyzing binding events in cultured cells, the ChIP assay can also be used to study protein–DNA or protein–protein interactions in primary cells or tissues, which can be very useful when studying events that may be tissue or developmental stage specific [6,7]. For example, in a recent analysis of promoters regulated by E2F1, the promoter for the plasminogen activator inhibitor (PAI-1) gene displayed different E2F binding patterns in different tissues: no E2Fs bound to the PAI 1 promoter in liver or kidney; E2Fs 3, 4, and 5 bound in the spleen, and only E2F1 bound in the testes [6]. As noted for the E2F family, many transcription factors are members of multiprotein families, each of which binds to the same consensus sequence. Therefore, redundancy can make determination of the role of a single family member difficult because creating a cell line or animal

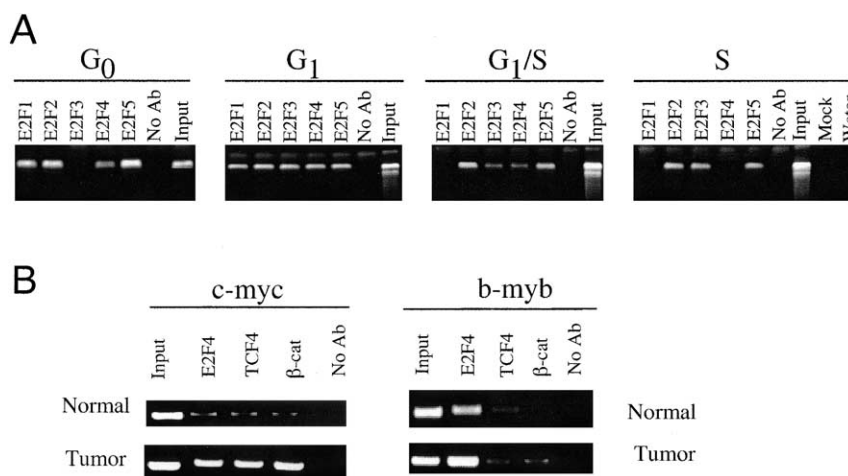


Fig. 4. Using ChIPs to study cell cycle- and tumorigenesis-dependent transcription factor binding. (A) Murine 3T3 cells were synchronized by serum starvation and stimulation and collected at the indicated points in the cell cycle and crosslinked. Chromatin was immunoprecipitated with antibodies against E2F1–5 or no antibody as a negative control. The immunoprecipitated chromatin was then amplified by 35 cycles of PCR with primers specific for the Rb promoter. The total, mock, and water samples were included as controls and are described in the legend to Fig. 3. (B) Human colon (normal) and colon tumor (tumor) tissue were crosslinked and immunoprecipitated with antibodies against E2F4, TCF4, and β -catenin. Immunoprecipitated DNA was amplified for 34 cycles of PCR with primers against the *c-myc* promoter (an E2F and a TCF4 target gene) and the *b-myb* promoter (an E2F, but not a TCF4, target gene). No Ab and input samples are included as controls and are as described in Fig. 3. Binding of E2F4, TCF4, and β -catenin to the *myc* promoter is enhanced in the tumor tissue whereas E2F4 binds to the *b-myb* promoter in both normal and tumor tissue.

that lacks one member may not result in a significant change in target gene expression. However, ChIP analysis of both wild-type and knockout cell lines or animals allows the investigator to examine how loss of one factor affects binding of another factor to target gene promoters. Comparison of the patterns of binding of the six E2F proteins to the thymidine kinase (TK), cyclin E, pRb, and dihydrofolate reductase (*dhfr*) promoters in wild-type and E2F1 null mouse embryos showed that in the absence of E2F1, E2F3 binding increases. Since expression of these genes is almost identical in the two populations of animals, these results suggest that E2F3 compensates for the loss of E2F1 in regulating expression of TK, cyclin E, pRb, and *dhfr* [6]. This adaptation of the ChIP technique for the analysis of DNA–protein interactions in embryos may be very helpful when studying nullizygous mice. For example, loss of many genes has been shown to result in embryonic lethality in mice. However, embryos can still be harvested and analyzed using the ChIP assay provided the embryos develop long enough to be identifiable.

Finally, many DNA-binding proteins play roles in tumor initiation and progression. ChIP analysis of tumors and nonneoplastic tissues can aid in understanding how binding of a factor of interest is altered during tumorigenesis. The TCF/LEF family of transcription factors interacts with β -catenin and activates target genes, leading to the development of colon tumors. A comparison of the binding patterns of the β -catenin/TCF4 transcriptional complex at the *c-myc* promoter (which was the first β -catenin target gene to be identi-

fied) and the *b-myb* promoter (which does not contain a TCF/LEF binding site and is included as a negative control) in normal human colon versus a colon tumor is shown in Fig. 4B. ChIP analysis of embryos and tissues, as opposed to cultured cells, does require a few modifications to the standard protocol; these modifications are described below.

2.4. What other factors are interacting with a protein when it is bound to the site of interest?

In addition to protein–DNA interactions, formaldehyde crosslinking can also be used to study protein–protein interactions. Immunoprecipitation after formaldehyde crosslinking, followed by Western blot analysis using an antibody against a second protein, may allow for the detection of some protein–protein interactions that are either too weak or too transient to be detected by traditional coimmunoprecipitation analyses. Fig. 5 shows a Western blot which indicates that DP1 can be immunoprecipitated after crosslinking using an antibody against E2F1. Although DP1 was certainly expected to be found in a protein complex with E2F1 (since these two proteins are known heterodimerization partners), this same assay should also allow the identification of new protein partners. Furthermore, ChIP combined with Western blot analysis may also allow for the identification of protein–protein interactions that occur only on DNA and not in solution. For example, DNA-bound protein complexes can be separated from non-DNA-bound fractions by loading the crosslinked

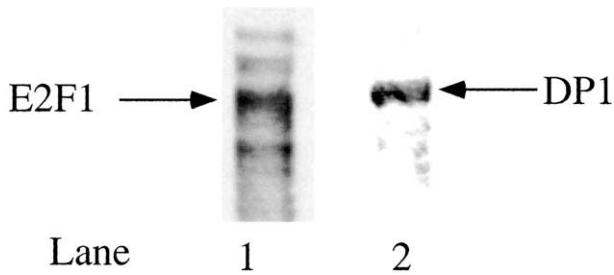


Fig. 5. ChIPs can be used to study protein–protein interactions. One gram of mouse liver tissue was crosslinked and immunoprecipitated with an antibody against E2F1. The crosslinks were reversed and the immunoprecipitated proteins were electrophoresed on a 10% polyacrylamide gel and transferred to a nitrocellulose membrane. The membrane was probed with antibodies against E2F1 (lane 1) and its heterodimerization partner DP1 (lane 2).

chromatin onto a CsCl step gradient before immunoprecipitation [7,8]. Free protein will equilibrate to a different position in the gradient than will protein bound to chromatin. Following 20 h of ultracentrifugation, fractions can be collected and a small amount of each fraction can be analyzed by electrophoresis to identify samples that contain DNA. Fractions containing crosslinked chromatin can be pooled and immunoprecipitated. Western blot analysis of these samples will detect protein–protein interactions that occur on DNA. A disadvantage of this approach is that it does not allow determination as to which DNA sequence was the site of the protein–protein interaction. However, once a positive interaction has been detected, a specific binding site can then be examined by standard ChIP analysis to determine if the identified protein–protein interaction is occurring at a DNA site of interest.

One problem with a combined ChIP–Western blot analysis is that the number of different protein–protein interactions that can be detected is limited by the relatively large amount of sample needed and the potentially large number of candidate antibodies to be tested. One method of amplifying the signal from rare proteins is to label cellular proteins with [³⁵S] methionine [9] prior to crosslinking. Alternatively, ChIP can be combined with mass spectrometry, resulting in a large-scale screen of protein–protein interactions. Such studies are underway at present and should provide a new and useful means of studying protein–protein interactions.

A novel, but yet untested, method for identifying all factors that bind to a promoter of interest combines transfected or episome-containing cell lines, ChIP, and mass spectrometry. First, a cell line must be created that harbors a plasmid or episome containing the promoter of interest having a GAL4 binding site inserted either upstream or downstream of the transcriptional start site. Next, these cells must be transfected with a vector expressing GAL4 protein, which should bind to the site within the promoter on the episome or plasmid. These

cells could then be crosslinked and DNA fragments containing the introduced promoter could be immunoprecipitated with an antibody against GAL4. After immunoprecipitation, all proteins bound to the Gal4 site-containing promoter could be eluted and identified by mass spectrometry. Although complicated in design, the major advantage of this technique is that proteins bound to the promoter of interest, and not to other cellular promoters, are specifically purified.

3. Methodology

A detailed standard ChIP protocol can be found in the article by Weinmann and Farnham [14]; the following are modifications to the standard ChIP protocol that must be made to further characterize *in vivo* binding sites, as described above.

3.1. Generation of cell lines containing constructs

To date, three different techniques have been used to create cell lines for ChIP analysis: transient transfection, stable integration, and introduction of episomes. Transient transfection is the least time consuming of the three techniques and both the calcium phosphate [10] and FuGENE (M.J. Oberley and S.R. Eberhardy, personal communication) methods have successfully been used in conjunction with ChIP. However, transfection with FuGENE can result in cells containing thousands of plasmids, requiring a reduction in the number of PCR cycles to be performed. For either transfection method, approximately 1.5×10^5 cells are transfected with 5 μ g of the plasmid containing the binding site and, if appropriate, 15 μ g of plasmid expressing an epitope-tagged wild-type or mutated version of the protein under study. However, the exact amount of plasmids to be transfected may need to be determined empirically. Cells are then crosslinked for the usual time at 24–36 h after transfection.

Stably transfected cell lines offer an advantage over transient transfection in that the introduced binding sites are integrated into the genome, possibly allowing for secondary DNA structure that more closely resembles the endogenous locus. First, 9 μ g of plasmid containing the DNA region of interest cloned upstream of a reporter gene (such as luciferase) and 1 μ g of plasmid containing a drug resistance marker are transfected into 1.5×10^5 cells using calcium phosphate. Next, cells are glycerol shocked and grown in selection medium containing the drug to which the presence of the second plasmid confers resistance. After 2 weeks of selection, individual clones are isolated and expanded. When working with stably transfected cell lines it is important to test several different clones to ensure that the site of integration does not affect protein binding or regulation of gene expression.

The third method of generating cell lines for ChIP analysis, introduction of episomes, is the most time consuming but also perhaps the most flexible. Unlike the plasmids used for transient transfection or stable integration, episomes can accommodate very large inserts of DNA (8–10 kb is not uncommon). Thus, regions of DNA can be studied in the context of several kilobase pairs of surrounding native sequence. Additionally, since episomes are capable of autonomous DNA replication, there is no concern about the effects of different sites of integration. Episomes typically contain a drug resistance marker, a multiple cloning site, and an origin of replication from Epstein–Barr virus, oriP. Cell lines are generated by transfecting episomes into Epstein–Barr virus-positive cells and growing the cells in medium containing the drug whose resistance gene is present on the episome. With both episome-containing and transiently transfected cell lines, however, caution must be taken not to create cells that contain more binding sites than molecules of transcription factors that bind to that site.

3.2. ChIP analysis of engineered cell lines

Very few modifications of the standard ChIP protocol are required for analysis of extrachromosomal or integrated constructs. However, the primers required for PCR analysis must be carefully chosen to amplify only the construct and not the endogenous loci. For cells containing plasmids, this can be ensured by designing one primer to hybridize to the inserted sequence under study and the other to hybridize to the plasmid backbone. Alternatively, the primer sequences can be chosen to hybridize to the junctions between plasmid vector and inserted sequence. In the case of cell lines containing stably integrated constructs, the constructs usually contain a reporter gene, such as luciferase. Therefore, one of the primers can be designed to hybridize to the reporter gene and the other primer to the sequence under study [5,11]. Analysis of introduced episomes, however, can be tricky due to the large size of most episomes. If the episome contains an altered sequence, such as a mutated, inserted, or deleted site, then one primer should be chosen to hybridize to this altered site. This approach was successfully used to analyze binding to an episome that contained the *myc* promoter in which either the E2F or the Me1a binding site was replaced with an *EcoRI* restriction site [12]. If it is not possible to design primers that hybridize to a sequence unique to the episome, the ChIP approach may still be useful since most cell lines contain multiple copies of the episome which can be visualized using significantly fewer cycles of PCR than needed for analysis of the endogenous locus. For the above mentioned cells containing episomes with sequences of the *myc* promoter (25–30 copies of the episome per cell), the number of PCR cycles was

decreased from 34 (which was needed to analyze the endogenous promoter) to 27 (for analysis of the episome).

3.3. Synchronization of cells

As mentioned above, cell cycle regulation of binding patterns can be analyzed by synchronizing cells before performing ChIP. A variety of synchronization techniques, including serum starvation and stimulation, drug treatment, and FACS, all work well in conjunction with ChIP. Therefore, the choice of synchronization method depends more on the cell type than any other factor. For example, some cell types, such as HeLa cells, do not efficiently stop cycling after the removal of serum, making serum starvation and stimulation a poor choice for this cell type. HeLa cells can, however, be efficiently synchronized at the beginning of S phase by the addition of 1 μg of aphidicolin per milliliter of medium for 24 h. Other cell types, such as Raji, are better synchronized by a thymidine/aphidicolin double block [13]. For all cell types tested to date in our laboratory, FACS is an excellent method to obtain synchronized populations of cells. We have found that the best method of sorting cells for ChIP analysis is to first crosslink whole cells in formaldehyde for 10 min and then stop the crosslinking reaction by adding glycine to a final concentration of 0.125 M and incubating for 5 min (see article by Weinmann and Farnham [14] for details). After swelling of the cells followed by centrifugation, the nuclei are then stained by the addition of approximately 50 μg of propidium iodide per 1×10^7 nuclei in a volume of 1 ml and incubated at 4 °C for 12 to 24 h. Following sorting, the crosslinked nuclei are then processed according to the standard ChIP protocol with no additional modifications, as the presence of propidium iodide does not interfere with subsequent steps. The beauty of crosslinking the cells before sorting is that it does not matter how long the sort takes because cells stop cycling once they have been fixed with formaldehyde.

3.4. ChIPs analysis of tissues, whole mouse embryos, or tumors

Prior to crosslinking tissues, whole mouse embryos, or tumors, the sample must first be minced into small pieces with a razor blade or scalpel. In general, 0.03 g of tissue per ChIP sample is sufficient but the exact amount required will depend on the abundance of the protein to be crosslinked, the immunoprecipitation efficiency of the antibody, and the crosslinking efficiency of the region of DNA to be examined. After mincing with a razor blade or scalpel, the tissue pieces are then placed into a 15- or 25-ml conical tube and approximately 10 ml of tissue culture medium or $1 \times$ PBS per every 0.3 g of tissue is added. Formaldehyde is added to a final concentration

of 1.0% and the samples are placed on a rotating platform at room temperature for 15 min, instead of the customary 10 min used for cultured cells. After 15 min, the crosslinking reaction is stopped by the addition of glycine to a final concentration of 0.125 M and incubation at room temperature for 5 min. The samples are then centrifuged at 200g to pellet the tissues pieces and the supernatant is removed. The tissue pieces are washed once with cold 1× PBS. A small volume of 1× PBS (1–2 ml per 0.3 g of tissue) is added and the tissue pieces are disaggregated either by 10 strokes in a Dounce homogenizer or by a 45-s pulse in a Medimachine (Becton–Dickinson). The resulting mixture is then centrifuged at 2000g, the supernatant is decanted, and the cell pellet is swelled in RSB buffer on ice for 15 min. From this point forward, the protocol is the same as for cultured cells (see article by Weinmann and Farnham [14]).

3.5. Western blot analysis

When samples are to be analyzed by Western blot analysis, instead of PCR, the only modification that must be made to the ChIP protocol is an alteration of the elution step. After washing the immunoprecipitates and removing the last traces of buffer, sodium dodecyl sulfate–polyacrylamide gel electrophoresis buffer is added directly to the pellet of Staph A cells. The samples are then boiled for 5 min and centrifuged at 2000g for 5 min and the supernatant is loaded onto the Western gel. Alternatively, the supernatant can be transferred to a clean tube and stored at –20 °C.

4. Concluding remarks

The ChIP assay has provided an exciting means for studying protein–DNA interactions in living cells. In this article we have described how the ChIP assay can be used, in place of gel shift analysis and *in vivo* footprinting, to identify precisely which nucleotides of a given DNA sequence are required for binding of a transcription factor or protein complex of interest and how the ChIP assay can reveal if binding patterns are cell cycle regulated. Additionally, we have described modifications necessary to use ChIP to study protein–DNA interactions in tissues or whole mouse embryos. This combination of ChIP and mouse genetics offers a unique ability to study the effects of genetic alterations

on transcriptional regulation of specific promoters. The ability to use tissues and embryos also allows analysis of transcription factor binding patterns during development, as well as in mouse and human tumor samples. At present, ChIP is also being combined with Western blotting and mass spectrometry to study protein–protein interactions. The ability to use the ChIP assay to study transcription factors *in vivo* may prove especially important in the analysis of tumors in which the transcription factor of interest may have altered DNA binding specificity and/or altered protein partners as compared to nonneoplastic tissue.

Acknowledgments

The authors thank Kathy Boyd for the numerous hours spent troubleshooting the ChIP protocol. We also thank Rob Chapman, Dirk Eick, Antonis Kirmizis, Scott Eberhardy, and Matt Oberley for sharing unpublished data and the entire Farnham laboratory for critical reading of the manuscript. Public Health Service Grants CA45250 (P.J.F.) and CA07175 (a NCI Cancer Center Core grant) were used to support this work.

References

- [1] A.S. Weinmann, S.M. Bartley, M.Q. Zhang, T. Zhang, P.J. Farnham, *Mol. Cell. Biol.* 21 (2001) 6820–6832.
- [2] A.S. Weinmann, P.S. Yan, M.J. Oberley, T.H.-M. Huang, P.J. Farnham, *Genes Dev.* (2002) 235–244.
- [3] S. Tommasi, G.P. Pfeifer, *Mol. Cell. Biol.* 15 (1995) 6901–6913.
- [4] M.F. Roussel, J.N. Davis, J.L. Cleveland, J. Ghysdael, S.W. Hiebert, *Oncogene* 9 (1994) 405–415.
- [5] J. Wells, K.E. Boyd, S.M. Bartley, P.J. Farnham, *Mol. Cell. Biol.* 20 (2000) 5797–5807.
- [6] J. Wells, C.R. Graveel, S.M. Bartley, S.J. Madore, P.J. Farnham, *Proc. Natl. Acad. Sci. USA* (2002), in press.
- [7] Z. Liang, M. Biggin, *Development* 125 (1998) 4471–4482.
- [8] B.S. Parekh, T. Maniatis, *Mol. Cell* 3 (1999) 125–129.
- [9] V. Botquin, H. Hess, G. Fuhrmann, C. Anastassiadis, M.K. Gross, G. Vriend, H.R. Scholer, *Genes Dev.* 12 (1998) 2073–2090.
- [10] R.X. Luo, A.A. Postigo, D.C. Dean, *Cell* 92 (1998) 463–473.
- [11] K.E. Boyd, P.J. Farnham, *Mol. Cell. Biol.* 19 (1999) 8389–8399.
- [12] T. Albert, J. Wells, J.O. Funk, A. Pullner, E.E. Raschke, G. Stelzer, M. Meisterernst, P.J. Farnham, D. Eick, *J. Biol. Chem.* 276 (2001) 20842–20890.
- [13] R.S. Maser, O.K. Mirzoeva, J. Wells, H. Olivares, B.R. Williams, R. Zinkel, P.J. Farnham, J.H.J. Petrini, *Mol. Cell. Biol.* 21 (2001) 6006–6016.
- [14] A.S. Weinmann, P.J. Farnham, *Methods* 26 (2002) 37–47.