Chromatin Immunoprecipitation for Determining the Association of Proteins with Specific Genomic Sequences In Vivo

Chromatin immunoprecipitation (ChIP) is a powerful and widely applied technique for detecting the association of individual proteins with specific genomic regions in vivo. In this technique, live cells are treated with formaldehyde to generate protein-protein and protein-DNA cross-links between molecules in close proximity on the chromatin template in vivo. A whole-cell extract is prepared, and the cross-linked chromatin is sheared by sonication to reduce average DNA fragment size to ~500 bp. The resulting material is immunoprecipitated with an antibody against a desired protein, modified (e.g., acetylated, phosphorylated, methylated) peptide, or epitope (in situations where the protein of interest is epitope-tagged). DNA sequences that directly or indirectly cross-link with a given protein (or modified variant) are selectively enriched in the immunoprecipitated sample. Thus, the method is not restricted to sequence-specific DNA-binding proteins. Reversal of the formaldehyde cross-linking by heating permits the recovery and quantitative analysis of the immunoprecipitated DNA. The amounts of specific genomic regions in control and immunoprecipitated samples are determined individually by quantitative PCR. The fold enrichment of certain chromosomal sequences (e.g., presumed binding sites) relative to other chromosomal sequences (e.g., presumed nonbinding sites) provides quantitative information about the relative level of association of a given protein with different genomic regions. Protein association with specific genomic regions can be performed under a variety of conditions (e.g., environmental change, cell-cycle status) and/or in wild-type versus mutant strains. Furthermore, as formaldehyde inactivates cellular enzymes essentially immediately upon addition to cells, ChIP provides snapshots of protein-protein and protein-DNA interactions at a particular time point, and hence is useful for kinetic analysis of events occurring on chromosomal sequences in vivo. In addition, ChIP can be combined with microarray technology to identify the location of specific proteins on a genome-wide basis (see Commentary). This unit describes the ChIP protocol for cells of the baker's yeast *Saccharomyces cerevisiae* (see Basic Protocol); however, it is also applicable to other organisms, although some organism-specific modifications related to cell lysis and sonication are necessary. A protocol for eluting immunoprecipitated protein-DNA complexes is also provided (see Alternate Protocol 1). As an alternative to gel electrophoretic analysis of the PCR products, a quantitative PCR analysis in real time with SYBR Green is also provided (see Alternate Protocol 2).

CHROMATIN IMMUNOPRECIPITATION

Materials

Saccharomyces cerevisae cells to be studied
37% formaldehyde: store up to 1 year at room temperature
2.5 M glycine, heat sterilized
TBS (*APPENDIX 2A*), ice cold
FA lysis buffer with and without 2 mM PMSF (see recipe), ice cold
ChIP elution buffer (see recipe)
20 mg/ml Pronase (Roche) in TBS; store up to 1 year at −20°C
TE buffer, pH 7.5 (*APPENDIX 2A*)
20 mg/ml DNase-free RNase A (see recipe)
10× loading buffer (see recipe)

Primary antibody against protein or epitope of interest
50% (v/v) protein A-Sepharose beads (Amersham Pharmacia Biotech) or
    equivalent in TBS
FA lysis buffer (see recipe), room temperature
FA lysis buffer (see recipe)/0.5 M NaCl
ChIP wash buffer (see recipe)
Primers (see Critical Parameters and Troubleshooting)
3000 Ci/mmol [^{32}P]dATP (optional; see annotation to step 30)

2-ml screw-cap microcentrifuge tubes with (relatively) flat bottoms
~0.5-mm-diameter silica-zirconia (BioSpec; preferred) or glass beads
Mini bead beater (BioSpec; preferred) or individual or multivortexer
5-ml syringe
15-ml conical tubes, disposable
25-G needles
Sonicator with microtip probe (e.g., Branson Sonifier 250)
End-over-end rotator
0.5-ml PCR tube
Spin-X centrifuge-tube filter (e.g., Corning)
65°C water bath
PCR-purification spin column (Qiagen)
Software for analyzing PCR primers and products

Additional reagents and equipment for growth of *Saccharomyces cerevisiae*
    cultures (*APPENDIX 3A*), phenol/chloroform extraction and ethanol precipitation
    (*APPENDIX 3A*), PCR (*APPENDIX 3F*), agarose gel electrophoresis (*APPENDIX 3A*), and
    nondenaturing acrylamide gel electrophoresis (*UNIT 6.5*)

CAUTION: When working with radioactive materials, take appropriate precautions to
avoid contamination of the experimenter and the surroundings. Carry out the experiment
and dispose of wastes in an appropriately designated area, following guidelines provided
by the local radiation safety officer (also see *APPENDIX 1D*).

Cross-link protein-DNA complexes in vivo

1. For each sample, grow 200 ml *Saccharomyces cerevisiae* to OD$_{600}$ = 0.6 to 0.8.

   CAUTION: *Keep cultures covered or work in a fume hood to avoid noxious formaldehyde
   fumes.*

   *The volumes of culture can be reduced (20 ml is a reasonable minimum) or increased
   depending on need. Typically, 20 to 40 ml yeast is used for an individual immunopre-
   cipitation, so the 200-ml volume permits multiple immunoprecipitations from the same
   cells. This is particularly useful for experiments involving the analysis of multiple factors
   or for carrying out independent immunoprecipitations involving the same factor for data
   reproducibility.*

2. Add 5.5 ml of 37% formaldehyde (1% final). Cross-link 15 to 20 min at room tem-
   perature by occasionally swirling flask or shaking slowly on a platform.

3. Add 30 ml heat-sterilized 2.5 M glycine and incubate an additional 5 min at room
   temperature.

   Glycine stops the cross-linking by reacting with formaldehyde.

Harvest cells

4. Centrifuge cells 5 min at 2500 × *g*, 4°C. Discard supernatant into a chemical waste
   container and resuspend pellet in 50 to 200 ml ice-cold TBS. Repeat once.

**Determining the
Association of
Proteins with
Specific Genomic
Sequences**

17.7.2

Supplement 23

Current Protocols in Cell Biology

5. Centrifuge cells for a third time 5 min at 2500 × g, 4°C. Discard supernatant and resuspend cells in 10 ml ice-cold FA lysis buffer.

6. Pellet cells by centrifuging in a benchtop centrifuge 5 min at 3000 rpm, 4°C. Discard supernatant.

   The cells can remain on ice for a few hours while other samples are being collected so that all samples may be processed as a group from this point onward. Alternatively, the cells may be frozen in liquid nitrogen or a dry ice/ethanol bath and stored up to several months at −80°C. This is particularly helpful if multiple samples are being generated during a time-course experiment. If cells are frozen, they must be thawed on ice before continuing with the procedure.

Lyse cells and isolate chromatin

For lysis using a mini bead beater (preferred)

7a. Resuspend the cell pellet in 1 ml ice-cold FA lysis buffer/2 mM PMSF. Fill three-quarters of a 2-ml flat-bottomed screw-cap microcentrifuge tube with ∼0.5-mm-diameter silica-zirconia or glass beads. Add cells, taking care to avoid introduction of bubbles, and screw the cap on tightly. Make sure there are no leaks.

   The mini bead beater is recommended because it is more efficient at breaking cells (multiple samples can be broken simultaneously). Silica-zirconia beads are more efficient at breaking cells than glass beads and are also recommended. To facilitate cell breakage with the mini bead beater, it is important that the final suspension nearly fill the tube. Do not break >160 OD$_{600}$ units of cells (i.e., <5 × 10^9 cells) in a single 2-ml tube; for larger cultures, split the cells into multiple tubes.

8a. Lyse cells 3 min with a mini bead beater at maximum speed. Remove sample and incubate 1 min in an ice-water bath. Repeat five times for a total breakage time of 18 min.

   This step assumes breakage with silica-zirconia beads. The cell breakage time with glass beads may be longer.

For lysis using an individual or multivortexer

7b. Resuspend in 250 µl FA lysis buffer/2 mM PMSF. Add 350 µl silica-zirconia or glass beads to a 2-ml microcentrifuge tube with relatively flat bottom. Add cells.

   When using a multivortexer (or standard vortexer), it is important to keep the volume small as this improves cell breakage. Do not break >160 OD$_{600}$ units of cells (i.e., <5 × 10^9 cells) in a single 2-ml tube; for larger cultures, split the cells into multiple tubes.

8b. Vortex continuously on an individual or multivortexer 30 min at full speed, 4°C.

   Success and reproducibility of the ChIP procedure is aided by complete (or near-complete) breakage of cells. In this regard, formaldehyde-cross-linked cells are considerably harder to break than untreated cells. The use of 1.5-ml microcentrifuge tubes with conical bottoms should be avoided because the narrow shape constricts bead movement, resulting in unequal lysis among samples. The 2-ml microcentrifuge tubes have a nearly flat bottom that allows the beads to vortex vigorously. The indicated vortexing or bead-beating conditions should be tested if a different device is used.

Isolate lysate

9. Cut a 5-ml syringe ∼1 cm below the flared opening (i.e., where the plunger is inserted) with a razor. Insert the smaller portion into a 15-ml disposable conical tube so that the flared portion of the truncated syringe rests on top of the conical tube opening, forming a microcentrifuge-tube holder.

10. Invert the sample tube and punch a hole in the bottom with a 25-G needle. Place the sample tube into the syringe/conical tube and punch a hole in the top cover with the same needle.

Macromolecular Interactions in Cells

17.7.3

11. Spin the assembly in a benchtop centrifuge 1 min at 1000 rpm, 4°C. Place the conical tube on ice. Discard the 2-ml centrifuge tube containing the dry beads after confirming the sample has been transferred to the 15-ml tube.

   Occasionally, beads will clog the pierced hole and prevent complete transfer of the sample. If this occurs, pierce the tube one or two more times and repeat the step in the same 15-ml tube. No additional buffer should be added.

Shear DNA

12. Transfer the sample to a standard 1.5-ml microcentrifuge tube. Microcentrifuge 15 min at maximum speed, 4°C. Discard the supernatant and add 1 ml ice-cold FA lysis buffer to the pellet.

   The pellet contains the cross-linked chromatin, cell debris, and unbroken cells. The purpose of this centrifugation step is to remove soluble protein, most of which is not cross-linked to DNA, as it might contribute to nonspecific background in the subsequent immunoprecipitations step. There is no need to resuspend the pellet at this point.

13. Holding the microtip probe near the bottom of the tube to prevent foaming, sonicate the sample 30 sec at 4°C using a continuous pulse at a power output of 20%. Cool in an ice-water bath >1 min. Repeat two more times.

   Take great care that the sample does not get too hot.

   If a different sonication device is used, empirically determine the conditions necessary to achieve the desired level of DNA shearing. The shear size is determined as described below (see Critical Parameters and Troubleshooting).

14. Microcentrifuge 30 min at maximum speed, 4°C. Transfer the supernatant to a fresh 15-ml disposable conical tube, add 4 ml ice-cold FA lysis buffer, and gently mix by inversion. Remove 250 µl for checking DNA fragment size and freeze the remaining chromatin solution in 800-µl aliquots in liquid nitrogen.

   Upon sonication, the cross-linked chromatin is solubilized and purified away from the pelleted material which contains cell debris and unbroken cells. The resulting chromatin solution constitutes the input sample for the subsequent immunoprecipitation. The frozen aliquots are stable for many months when stored at −70°C and are suitable for immunoprecipitations.

Check chromatin-fragment size

15. Add 250 µl ChIP elution buffer and 20 µl of 20 mg/ml Pronase in TBS to the 250-µl chromatin aliquot. Incubate 2 hr at 42°C, followed by 6 hr at 65°C. Phenol extract and ethanol precipitate sample (*APPENDIX 3A*).

   While it is convenient to perform the reaction in a PCR machine overnight, it could just as easily be done in heat blocks or water baths. The same is true of the incubation described in step 26.

16. Resuspend in 30 µl TE buffer, pH 7.5, add 1 µl of 20 mg/ml DNase-free RNase A, and incubate 15 min at 37°C. Add 3 µl of 10× loading buffer and electrophoretically separate the material on a 1.5% agarose gel (*APPENDIX 3A*).

   Fragments should be between 100 to 1000 bp, with an average length of 400 to 500 bp.

   It is important to shear DNA fragments down to an average length of 400 to 500 bp. Longer fragments will increase the background and will decrease the resolution of the region to which the protein associates (see Commentary).

Immunoprecipitate

17. Incubate 800 µl chromatin solution with 10 µl primary antibody against the protein or epitope of interest and 20 µl of 50% (v/v) protein A–Sepharose beads in TBS on an end-over-end rotator 90 min at room temperature.

The actual amount of antibody needed has to be empirically determined and can vary considerably. The idea is to have an excess of antibody to efficiently precipitate at least 50% of the antigen in question. One way to assess the efficiency of antigen immunoprecipitation is to determine the amount of antigen present in the sample before and after the immunoprecipitation. An aliquot of 30 µl chromatin solution, taken before and after immunoprecipitation, is usually sufficient to visualize the protein of interest via immunoblotting (UNIT 6.2) and standard chemiluminescent detection; however, the samples have to be boiled in SDS/PAGE sample buffer for 30 min prior to loading in order to reverse the formaldehyde cross-links. The immunoprecipitation conditions can be varied (e.g., time, temperature, salt concentration, presence of detergents) if necessary.

Protein A–Sepharose beads are used here because they work well with most monoclonal and polyclonal sera derived from mouse and rabbit, respectively. In some cases, the use of other beads (e.g., protein G–Sepharose) may improve binding of some antibodies, including rat IgG (see Table 7.2.1).

18. Microcentrifuge beads 1 min at 3000 rpm, room temperature. Transfer 300 µl supernatant into a 0.5-ml PCR tube labeled "INPUT." Discard the rest of the liquid.

Wash beads

19. Resuspend beads in 700 µl FA lysis buffer, room temperature, and transfer mixture into a Spin-X centrifuge-tube filter.

    The use of Spin-X filters aids in the recovery of the beads after washes and results in better uniformity between different samples. The procedure is also substantially faster with the filters, particularly when multiple samples are processed simultaneously. Alternatively, one could use conventional microcentrifuge tubes for the washes and aspirate the supernatant with a narrow-bore pipet tip after each spin.

20. Place the filter into a 1.5-ml microcentrifuge tube and mix sample 3 min on an end-over end rotator. Microcentrifuge 2 min at 3000 rpm, room temperature. Discard the flow-through liquid at the bottom of the tube.

21. Add 700 µl FA lysis buffer, room temperature, to the beads and repeat step 20.

Elute protein from beads

22. Wash beads for 3 min each with 700 µl FA lysis buffer/0.5 M NaCl, 700 µl ChIP wash buffer, and finally 700 µl TE.

    For many polyclonal antibodies, the more stringent washes in this step result in a cleaner signal, while gentle washes frequently lead to an unacceptably high background. For some antibodies (e.g., monoclonal against peptide epitopes; see Alternate Protocol 1), repeated washes with FA lysis buffer, which are gentler, might be more appropriate.

23. Place filter unit containing the beads into a new 1.5-ml microcentrifuge tube and add 100 µl of ChIP elution buffer. Gently pipet up and down two or three times in order to dislodge beads from the filter. Incubate 10 min in a 65°C water bath.

    A water bath is used instead of other heating apparatuses in order to improve heat transfer.

24. Microcentrifuge beads 2 min at 3000 rpm, room temperature. Discard filter with beads. Transfer the eluate into a 0.5-ml PCR tube labeled "IP."

Reverse cross-links and purify DNA

25. Add 80 µl TE and 20 µl Pronase in TBS to the IP tube. Combine 20 µl INPUT material (step 18), 100 µl ChIP elution buffer, 60 µl TE, and 20 µl TBS into a new 0.5-ml PCR tube.

26. To reverse cross-links, place tubes into a PCR machine. Incubate 2 hr at 42°C, followed by 6 hr at 65°C. Store samples at 4°C until use.

Macromolecular Interactions in Cells

17.7.5

The incubation at 42°C allows for Pronase digestion of cross-linked polypeptides, while the 65°C incubation results in a reversal of the formaldehyde cross-links.

27. Purify DNA using a Qiagen PCR-purification spin column as per manufacturer's instructions.

    This will require double loading of the spin column (i.e., 600 μl spin through and then repeat).

    Alternatively, add 20 μl of 4 M LiCl and purify by extracting with 25:24:1 phenol/chloroform/isoamyl alcohol, followed by extraction with chloroform and ethanol precipitation (APPENDIX 3A). It is useful to add 2 μl of Pellet Paint (Novagen) prior to the addition of ethanol, as this aids both the ethanol precipitation and visualization of the very small pellet.

28. Resuspend in 300 μl TE and store at −20°C.

    DNA pellets stored in this fashion should be stable for years.

Perform quantitative PCR

29. Design primer pairs for the desired genomic regions to be examined.

    Success in obtaining high-quality data is critically dependent on good primer design (see Critical Parameters and Troubleshooting). In general, primers should be 20 to 30 bases long with a T_m of 55° to 60°C. The design of good primers is greatly facilitated by commercially available software packages such as Oligo 6.6 (see http://www.oligo.net) or Primer Express 1.5 (see http://www.appliedbiosystems.com). Most primers require no purification or special treatment prior to PCR. Amplification products should be 75 to 350 bp; longer products should be avoided, as the amplification efficiency is substantially lower. A final primer concentration of 1 μM works well for most primers, but in some instances, improved product specificity may be obtained by lowering the final primer concentration 5 to 10 fold. Refer to APPENDIX 3F for more information.

30. Dilute INPUT DNA (obtained from step 18) in three separate tubes by a factor of 5, 10, and 20. Set up standard PCR reactions (APPENDIX 3F) with 2 μl DNA sample, primers at 1 pmol/μl, and total reaction volumes of 10 to 50 μl. If PCR products will be detected by radioactivity, add 1 μCi of 3000 Ci/mmol [^{32}P]dATP.

    For a typical measurement, the three dilutions of input DNA are tested along with duplicate immunoprecipitated samples (or undiluted and 5-fold diluted immunoprecipitated samples). This permits an assessment of whether the assay is being performed in the linear range as well as of the reproducibility of the PCR reaction. The immunoprecipitated DNA is typically used without dilution, although it is useful to analyze different amounts to ensure that it is also in the linear range.

    There are several key parameters for achieving an optimum PCR reaction. For example, it is very important to have a quality repeat pipettor that can reproducibly dispense 2 μl DNA. Pipetting inaccuracies at this stage will lead to greater well-to-well variability and poorer reproducibility among identical samples. Additionally, multiple primer pairs (up to 4 to 5) can be included in the same reaction, provided that the PCR products can be unambiguously resolved from each other by gel electrophoresis. This permits simultaneous and internally controlled analysis of multiple genomic regions in a single reaction. However, it is critical to ensure that there is no competition between the different primer pairs and PCR products. Also, comparable results are obtained when PCR reactions are performed in volumes between 10 to 50 μl; using smaller volumes reduces the cost and facilitates loading of the reaction products on gels. See Critical Parameters and Troubleshooting for a discussion of primer choice.

    Detection of PCR products by [^{32}P]label is recommended over detection by ethidium bromide or SYBR Green (see Alternate Protocol 2) staining as it improves the sensitivity and extends the linear range of detection; however, it necessitates using the usual precautions in working with radioactivity.

31. Carry out hot-start PCR using the following thermal cycling parameters.

Initial step:	10 min	95°C	(denaturation)
26 cycles:	30 sec	95°C	(denaturation)
	30 sec	55°C	(annealing)
	1 min	72°C	(extension)
Final step:	4 min	72°C	(final extension).

These conditions are generally appropriate for most situations. The annealing temperature may have to be adjusted if the melting temperatures of the primers is substantially above or below 55°C. The number of cycles might also have to be adjusted in some cases if reactions are not in the linear range. See Critical Parameters and Troubleshooting for more details.

Analyze PCR products

32. Add the appropriate loading buffer to the PCR products, and analyze by electrophoresis on nondenaturing polyacrylamide (*UNIT 6.5*) or agarose gels (*APPENDIX 3A*).

    The gels should be stained either with ethidium bromide or SYBR Green dyes, or analyzed by autoradiography or PhosphorImager.

33. Quantitate the relative amount of PCR products using appropriate software for the accompanying instrument.

34. Calculate the apparent immunoprecipitation efficiency for a specific fragment by dividing the amount of PCR product obtained with the immunoprecipitated DNA by the amount obtained with the input DNA.

    A volume of 2 μl immunoprecipitated DNA sample (1/150 total immunoprecipitated material) contains ∼200 times the number of cell equivalents as 2 μl INPUT sample that has been diluted 5-fold (1/30,000 of the original aliquot that was immunoprecipitated). Thus, if the amount of PCR product in the immunoprecipitated sample is equal to the amount of PCR product in the 5-fold diluted INPUT sample, the apparent immunoprecipitation efficiency is 0.5%. The apparent immunoprecipitation efficiency for the background signal is typically ∼0.025% to 0.05%, and it should not be higher than 0.1%.

SPECIFIC PEPTIDE ELUTION OF PROTEIN-DNA COMPLEXES IMMUNOPRECIPITATED FROM CROSS-LINKED CHROMATIN

Peptide elution represents an alternative method for removal of immunoprecipitated protein-DNA complexes from beads. In this procedure, beads containing the immunoprecipitated complexes are incubated with high concentrations of a peptide recognized by the antibody used in the immunoprecipitation. The added peptide competes with the protein antigen of interest for binding to the antibody, and specifically liberates the protein-DNA complexes from the beads. The high specificity of peptide elution reduces the nonspecific background (typically by a factor of 2 to 4), which makes it the method of choice, particularly for applications where the expected immunoprecipitation signal is low. Peptide elution is especially useful for chromatin immunoprecipitation experiments involving proteins that are tagged with the HA or myc epitopes (in single or multiple copies); however, it would also be appropriate in cases where the antibody used for the immunoprecipitation was generated against a defined peptide sequence. Peptide elution is slightly more expensive than conventional elution, due to the cost of the peptide. In general, peptide elution should be used in conjunction with gentle washes during the immunoprecipitation procedure described below, which minimizes antigen leaching. Stringent washes, such as those employed in the main method (see Basic Protocol), will often result in signals that are several-fold lower, with little or no improvement in background. Finally, peptide elution may vary in quality depending on factors such as the number of epitopes in the antigen and the relative stability of the antibody-antigen interaction.

Additional Materials (also see Basic Protocol)

1 mg/ml peptide (e.g., myc, HA) in TBS (see *APPENDIX 2A* for buffer)

For this protocol, follow steps 1 to 21 of the main method (see Basic Protocol), replace steps 22 to 25 with the following, and continue with step 26 onwards.

22. Repeat FA lysis buffer wash (see Basic Protocol, steps 20 and 21) three additional times for a total of five washes.

> *Repeated washes with FA lysis buffer are much more gentle than the single washes with FA lysis buffer/0.5 M NaCl, ChIP wash buffer, and TE used in the Basic Protocol and result in higher signal-to-background ratios.*

23. Place the Spin-X centrifuge-tube filter unit containing the beads into a new 1.5-ml microcentrifuge tube and add 100 μl of 1 mg/ml peptide (typically myc or HA) in TBS. Gently pipet up and down two or three times in order to dislodge beads from the filter. Incubate 30 min at 30°C.

24. Microcentrifuge beads 2 min at 3000, room temperature. Discard filter with beads. Transfer the eluate into a 0.5-ml PCR tube suitable for PCR labeled "IP."

25. Add 150 μl TE, pH 7.5, 250 μl of ChIP elution buffer, and 20 μl of 20 mg/ml Pronase in TBS.

ALTERNATE PROTOCOL 2

ANALYSIS OF CHROMATIN IMMUNOPRECIPITATION EXPERIMENTS BY REAL-TIME QUANTITATIVE PCR WITH SYBR GREEN

Quantitative PCR (QPCR) analysis in real time with SYBR Green has several advantages over the analysis of PCR reactions by gel electrophoresis (see Basic Protocol, step 32). First, the method saves considerable time because no gels are involved and because quantitative values are obtained directly from the data curves and do not require densitometry or phosphor imager analysis. As a consequence, this approach permits very rapid analysis of much larger numbers of chromatin immunoprecipitation samples than can be performed with the Basic Protocol. Using standard 96-well instruments, it is a straightforward procedure to analyze 100 to 200 samples/day (in replicates of three) with only 1 to 2 hr of total setup time. With newer 384-well instruments and automated robotics equipment, sample throughput can be further increased to thousands per week. Second, the data generated by this procedure are more accurate and reproducible, because quantitative values are determined from continuous sampling throughout the PCR reaction rather than a single end-point determination. Furthermore, the quality and "linear range" of every PCR reaction are directly visualized. Third, the procedure is significantly safer for the researcher, as no radioactive materials or toxic acrylamide are used. The major disadvantage of this procedure is that the measurements are performed individually and hence are not internally controlled, whereas the Basic Protocol permits the simultaneous analysis of multiple genomic regions in a single PCR reaction (provided the individual primer pairs function independently). As such, the Basic Protocol is more useful for analyzing the same small set of genomic regions under multiple experimental conditions and for simultaneous analysis of electrophoretically distinguishable alleles of a given genomic region.

SYBR Green is a sensitive and highly selective double-stranded DNA (dsDNA)–binding dye that remains associated even at the high temperatures normally used for PCR template extension. Real-time PCR reactions involving SYBR Green are performed with standard oligonucleotide primers, and hence are much less expensive than real-time PCR reactions using fluorophore-conjugated oligonucleotides (e.g., TaqMan or Lux probes). Measurements of SYBR Green fluorescence at the polymerase extension step of PCR, when

Determining the Association of Proteins with Specific Genomic Sequences

17.7.8

plotted against PCR cycle number, provide both a qualitative assessment of the progress of the PCR and a way to quantitate the relative amount of DNA template initially present in the reaction. Typical real-time QPCR graphs feature the plot of the $\log_{10}$(Net fluorescence) on the y axis versus the PCR cycle number on the x axis, and usually contain three well-defined stages: (1) baseline, (2) linear, and (3) plateau. In the baseline stage, the amount of DNA product formed is still below the sensitivity threshold of SYBR Green, so product formation is undetectable. This part of the curve is typically used as a baseline for SYBR Green signal drift. The linear part of the curve is the most important from the analytical standpoint, because it is at this stage that the rate of PCR product accumulation is both constant on a per-cycle basis and readily detectable by increased SYBR Green fluorescence. Finally, as all of the SYBR Green in the reaction becomes bound to the recently synthesized PCR products, the amount of fluorescence stays constant from cycle to cycle and the reaction reaches a plateau.

In the protocol described below, PCR is performed under special conditions that minimize the inhibitory effects of SYBR Green on *Taq* activity and maximize the linear range of product detection. After amplification is complete, raw data are stripped of outliers and exported in a format readable by a spreadsheet program such as Microsoft Excel. Finally, data points from replicate samples are averaged, and mean values are further manipulated and ultimately compared to some internal reference or control.

Additional Materials (also see Basic Protocol)

> Input DNA (see Basic Protocol, step 28)
> Immunoprecipitated fragments ("IP" sample; see Basic Protocol, step 23)
> 2× SYBR Green *Taq* mix (see recipe)
>
> Real-time PCR machine and corresponding software (e.g., ABI)
> 96-well PCR plates (ABI, cat. no. 4306737) and optical adhesive covers
> Centrifuge with swinging-bucket rotor and microtiter plate adapter
> Spreadsheet program (e.g., Microsoft Excel)

Set up PCR reactions

1. Dilute input DNA to an approximate equivalent of 1×10^6 cells/ml in TE buffer, pH 7.5.

   If immunoprecipitations were performed as described in the Basic Protocol, then a 1:25 dilution of the input sample from step 28 will result in 1:1000 overall dilution and will correspond to ~5×10^8 to 1×10^9 cell equivalents.

2. If necessary, resuspend immunoprecipitated fragments in TE buffer, pH 7.5, so that the approximate cell equivalent is 1×10^9 cells/ml.

   Immunoprecipitated DNA derived from the IP sample obtained by the Basic Protocol (step 23) is appropriately diluted and needs no further treatment.

3. Prepare PCR primer stocks by mixing each primer pair at a final concentration of 3.3 μM in TE buffer, pH 7.5.

   It is critical to test newly obtained primer pairs for amplification specificity and performance under conditions that will be used for real-time PCR with SYBR Green (see Critical Parameters and Troubleshooting). SYBR Green can inhibit PCR reactions, and primer pairs that are appropriate for quantitative PCR analysis in the absence of SYBR Green may not work well in the presence of SYBR Green. High-quality primer pairs should result in ~1.9-fold amplification/cycle (this can be determined from quantitative analysis of raw fluorescence data for each cycle, which is generally available on commercial instruments). Amplified material at the completion of the PCR should contain only one band

(as assayed on high-percentage agarose or polyacrylamide gels). Specificity information may also be obtained by running dissociation curves on reactions following the conclusion of the PCR run. Typically, samples are melted for 15 min at 95°C, cooled to 60°C, and then slowly heated back up to 95°C over a period of 20 min. Plotting the first derivative of the fluorescence against the temperature allows for simple visual identification of sample heterogeneity. Some instrument-specific software packages have built-in modules for dissociation curve analysis.

4. Select and label the wells to be used in the run.

   In general, individual samples should be run in triplicate. Obvious outliers occur with some frequency, generally at <5%. Triplicate analysis of samples permits removal of those outliers while still allowing for inclusion of two accurate measurements for each sample. While this reduces the number of different samples that can be run at any given time, the resulting data is much more reliable and accurate.

   For each primer pair examined, the input DNA samples should be run alongside the immunoprecipitated samples. Amplification efficiencies among different primer pairs vary slightly on a per-cycle basis, but those slight variations in efficiency translate into substantially different amounts of amplified material in the cycle range used for analysis. Precise quantitation of relative binding cannot be accurately performed without primer pair–specific input signal.

   Detailed instructions on the use of the real-time PCR machine and general issues, e.g., calibration and camera exposure settings, are addressed in the documentation that accompanies the instruments.

5. Program the real-time PCR machine as follows:

1 cycle:	10 min	95°C	(initial denaturation)
40 cycles:	30 sec	95°C	(denaturation)
	30 sec	53°C	(annealing)
	30 sec	72°C	(extension).

   Collect the data only at 72°C.

   The annealing temperature may have to be adjusted if the melting temperatures of the primers are substantially above or below 53°C. If the desired amplification product is >500 bp (this is not recommended), the extension time at 72°C should be increased to 1 min. See Critical Parameters and Troubleshooting for more details.

6. Using a small-volume automatic pipettor (20-µl capacity), place a 2-µl aliquot of each DNA template into the appropriate wells of a 96-well PCR plate. Gently tap the plate to allow the sample droplets to fall to the bottoms of the wells.

   It is very important to have a quality repeating pipettor that can reproducibly dispense small volumes of sample into the wells. Pipetting inaccuracies at this stage will lead to greater well-to-well variability and poorer reproducibility among identical samples.

7. Using a small-volume automatic pipettor (20-µl capacity), place a 3-µl aliquot of primer mix (see step 3) into the relevant wells and tap the plate a few times to settle the contents.

   On many real-time PCR machines, results from 10-µl reactions are virtually indistinguishable from those of 25- and 50-µl reactions in their accuracy and reproducibility. The use of 10-µl reactions provides substantial savings in reagent costs. On some machines, the minimal reaction volume needed for accurate and reproducible results may be greater.

8. With a larger automatic pipettor (100-µl capacity), add 5 µl of 2× SYBR Green *Taq* Mix to every assayed well. Place microtiter plate into appropriate microtiter plate adapter and centrifuge 1 min at 200 × *g*, room temperature, in a swinging-bucket rotor.

The 2× SYBR Green Taq Mix contains a variety of components that considerably reduce the inhibitory effect of SYBR Green, thereby resulting in more reproducible signals that require fewer amplification cycles. Comparable mixes containing proprietary buffers can be obtained commercially. It is critical that quantitative PCR reactions containing SYBR Green be performed under conditions of efficient amplification (e.g., 1.9-fold amplification/cycle).

9. Seal plate with clear optical adhesive covers, overlay foam compression pad with gold side facing up, and place into the real-time PCR machine. Secure lid.

   The details of this step may differ, depending on the machine.

10. Start the PCR protocol (see step 5). After completion, save the run for future analysis.

DATA analysis

11. Open the file containing the real-time data according to the manufacturer's instructions for the instrument.

    Although the specific protocol will depend on the software and instrument, the overall logic and approach to the analysis of real-time data is generally applicable.

12. Look at the different curves and set the baseline as needed.

    Generally, the baseline should be set from cycle 3 to the cycle just prior to where the curves start increasing in a linear fashion. It is desirable to have at least 10 cycles for the calculation of the baseline, as this results in increased accuracy in the subsequent calculations of the threshold cycle.

13. Change the value in the Threshold box to be about halfway up in the linear range, and apply changes to the data set.

    The threshold cycle is defined to be the PCR cycle at which the fluorescence is 10 times (10 is the default multiplier) the standard deviation obtained in the baseline calculation. When the multiplier is set to 10, the fluorescence at the threshold cycle is considered the lowest fluorescence value that is significantly above the background. In practice, this number frequently lies in the nonlinear range of many of the curves. For later calculations, it is easier to manually set the fluorescence value used to calculate the threshold cycles to 0.04. At this value, all the curves should be in the linear range and well above the baseline, allowing for far more accurate comparisons of the threshold cycles. On occasion, however, it will be necessary to adjust this value either up or down to better reflect the linear range of net fluorescence for most of the curves.

14. Manually select one group of triplicates and visually inspect their amplification plots. If curves are essentially superimposable and the threshold cycle (C_T) values are close to each other (maximal and minimal replicates within 1 cycle, preferably within 0.5 cycles), proceed to the next triplicate sample. Otherwise, remove the outlier and continue to the next triplicate.

    Decisions regarding the removal of some outliers could either be straightforward or judgment calls, depending on the circumstances. In cases where two out of three curves are superimposable while the third is clearly off by more than a cycle, it is a fairly easy decision to consider the third replicate an outlier. If the curves are closer, the decision on which one to eliminate, if any, becomes much more difficult. As a general guide only, if the spread between the lowest and highest C_T values is less than 0.5, it is probably safe to average all the C_T values (see step 12). If the C_T range is <1 but >0.5, the data are less reliable and the decision to remove any data points should probably be made on a case-by-case basis. It is highly recommended that the PCR be repeated for samples where the C_T ranges are >1 with no two curves superimposable.

15. Proceed to analyze the data for all triplicates in the manner described above. Save the results in a different file.

16. Export the data to a spreadsheet program such as Microsoft Excel by using built-in filters. The file should not contain omitted wells (see step 14) and should be in a column format containing well positions, descriptors, and C_T values for each selected well.

Final calculations are most easily handled in a spreadsheet but could also be performed with a scientific calculator.

17. Open the exported file. Proceed to average triplicate measurements for each sample in a new column ($^{AVERAGE}C_T$).

For some samples, there may be only two measurements left as a result of the removal of outliers in step 14.

18. For each primer pair, subtract the $^{AVERAGE}C_T$ (INPUT) from $^{AVERAGE}C_T$ (IP) in a new column. This number is the $^{NET}C_T$.

This value represents the difference in cycles between the immunoprecipitated sample and the input DNA.

19. Subtract $^{NET}C_T$ for one primer pair (experimental; EXPT) from the $^{NET}C_T$ of another primer pair that serves as a reference or a control (CTRL) in a new column. The resulting value is $^{NET}C_T^{EXPT}$ − CTRL. Repeat $^{NET}C_T$ subtraction of control primer for all other experimental primers.

It is very desirable to have a control primer pair that can be used to assess the relative cross-linking efficiency at promoters of interest. Frequently, the control primer pair could be specific for a DNA region that does not bind to the immunoprecipitated protein of interest. The signal from the control primer pair could then be considered the background, and the binding efficiency of the protein to different promoter regions could be expressed as fold over background binding.

20. Evaluate the expression POWER ([mean primer slope], [− $^{NET}C_T^{EXPT}$ − CTRL]), where the [mean primer slope] is the base and [− $^{NET}C_T^{EXPT}$ − CTRL] is the exponent. Repeat the process with other primers by using the different $^{NET}C_T^{EXPT}$ − CTRL values calculated in step 19.

The actual value calculated in the POWER expression above is the degree of occupancy of the immunoprecipitated protein at the sequence of interest relative to that of a control (or background) DNA region.

Perfectly efficient PCR, in which the number of amplified molecules doubles every cycle, has a primer slope of 2. As defined, this value is independent of primer pair sequence, target sequence length, and other variables that under normal circumstances may adversely affect the efficiency of amplification. In practice, however, the mean primer slope is almost always <2 and is slightly variable from primer pair to primer pair, mostly due to differences in primer sequence and other parameters (e.g. GC content and length of amplified sequence). For the majority of primers designed to amplify S. cerevisiae promoter sequences, the mean slope is 1.9 ± 0.06, and this value can be safely used in the calculations above. However, it is still good practice to calculate representative slopes for every newly synthesized primer pair; any substantial deviation from 1.9 (±0.06), especially to the downside, should be viewed suspiciously. Frequently, a slope that substantially deviates from the normal range is indicative of problems in the amplification.

The mean primer slope for a given primer pair is most easily calculated by performing linear regression on the linear portion of the amplification plot (log net fluorescence versus PCR cycle). In order to perform this calculation, it is first necessary (if possible) to export a file containing the fluorescence values for all the wells in use at every PCR cycle (see software manuals for more information). It is then possible to use linear regression to estimate the slope within the linear range (i.e., by using the MS Excel function LINEST). Since it is rather time-consuming to calculate slopes for an entire 96-well plate one-by-one, it may well be worthwhile to write a macro (or a stand-alone program) to automate this process.

REAGENTS AND SOLUTIONS

Use deionized or distilled water in all recipes and protocol steps. For common stock solutions, see **APPENDIX 2A**; *for suppliers, see* **SUPPLIERS APPENDIX**.

ChIP elution buffer

50 mM Tris·Cl, pH 7.5 (*APPENDIX 2A*)
10 mM EDTA (*APPENDIX 2A*)
1% (w/v) SDS (*APPENDIX 2A*)
Store up to 1 year at room temperature

ChIP wash buffer

10 mM Tris·Cl pH 8.0 (*APPENDIX 2A*)
0.25 M LiCl
1 mM EDTA (*APPENDIX 2A*)
0.5% (v/v) Nonidet P-40
0.5% (w/v) sodium deoxycholate
Store up to 1 year at room temperature

FA lysis buffer with and without 2 mM PMSF or 0.5 M NaCl

For FA lysis buffer:
50 mM HEPES: adjust pH to 7.5 with KOH
150 mM NaCl
1 mM EDTA (*APPENDIX 2A*)
1% (v/v) Triton X-100
0.1% (w/v) sodium deoxycholate
0.1% (w/v) SDS
Store up to 1 year at room temperature

For FA lysis buffer/2 mM PMSF add 100 mM phenylmethylsulfonyl fluoride (PMSF) in ethanol (store up to 1 year at −80°C) to a final concentration of 2 mM just before use. For FA lysis buffer/0.5 M NaCl, change the amount of NaCl added to 0.5 M. Store up to 1 year at room temperature

Loading buffer, 10×

20% (w/v) Ficoll 400
0.1 M disodium EDTA, pH 8 (*APPENDIX 2A*)
1.0% (w/v) sodium dodecyl sulfate
0.25% (w/v) bromphenol blue
0.25% (w/v) xylene cyanol (optional; runs ~50% as fast as bromphenol blue and can interfere with visualization of bands of moderate molecular weight, but can be helpful for monitoring very long runs)

RNase A stock solution (DNase-free), 2 mg/ml

Dissolve RNase A (e.g., Sigma) in DEPC-treated H_2O (*APPENDIX 2A*) to 2 mg/ml. Boil 10 min in a 100°C water bath. Store ≤1 year at 4°C.

SYBR Green Taq mix, 2×

12 mM Tris·Cl, pH 8.3 (*APPENDIX 2A*)
50 mM KCl
5 mM $MgCl_2$

**Macromolecular
Interactions in
Cells**

17.7.13

150 mM trehalose (Sigma)

100 mM betaine (Aldrich)

0.2% (v/v) Surfact-Amps 20 (Pierce; active ingredient, Tween 20)

0.2 mg/ml nonacetylated BSA (Sigma, B8667)

1 μM 5(6)-carboxy-X-rhodamine (ROX; Helix Research)

0.133× SYBR Green (Molecular Probes; final dilution 1:75,000)

Store solution with above components indefinitely at −80°C or up to several
  months at 4°C in the dark

Add the following immediately prior to use

0.5 mM each dATP, dTTP, dCTP, dGTP

0.2 U/μl hot-start *Taq* DNA polymerase

*The amount of Taq DNA polymerase may need to be slightly adjusted to account
for batch/activity variations among different manufacturers.*

COMMENTARY

Background Information

Direct protein-DNA contacts and indirect protein-DNA interactions regulate fundamental chromosomal functions such as DNA replication, gene expression, and chromosome segregation. Thus, knowledge about the distribution of particular proteins on specific chromosomal DNA sequences can provide important insights into the mechanisms that govern chromosomal functions, structure, and organization. In vivo footprinting methods provide high-resolution mapping of protein-DNA interactions but cannot directly identify the chromatin-associated protein(s) responsible for the footprint. Chromatin immunoprecipitation, by contrast, is ideally suited for determining the identity of proteins associated with specific DNA sequences in vivo, albeit with lower resolution (≤1 kbp).

Two groups, Gilmour and Lis and Solomon and Varshavsky, independently pioneered cross-linking and immunoprecipitation methods for in vivo chromatin analysis (Gilmour and Lis, 1984; Solomon and Varshavsky, 1985; Solomon et al., 1988; Gilmour et al., 1991). These methods exploited cross-linking to conserve in vivo chromatin structures and permit their isolation under the stringent conditions necessary to isolate soluble chromatin. Cross-linked protein-DNA complexes were purified by cesium chloride centrifugation (a time-consuming step) and subjected to immunoprecipitation. Their methods were distinguished primarily by the cross-linking agent: Gilmour and Lis employed UV irradiation, while Solomon and Varshavsky used formaldehyde. The biochemical characteristics of each method have been discussed extensively (Gilmour and Lis, 1984; Solomon and Varshavsky, 1985; Orlando et al., 1997,

and references therein). In short, UV irradiation cross-links only protein-DNA complexes in direct contact, which limits its use. Formaldehyde reacts with primary amines on amino acids and DNA and RNA bases, reversibly forming a covalent adduct between two primary amines in close proximity to each other (≤2 Å). Because protein-protein adducts are formed in addition to protein-DNA adducts, chromatin-associated proteins not directly bound to DNA can be cross-linked to DNA via other proteins such as histones, significantly broadening the applicability of this procedure. Cross-linking with formaldehyde is also more easily accomplished than UV irradiation, especially with larger culture volumes.

Formaldehyde-based chromatin immunoprecipitation was simplified and adapted for use in other experimental systems, including budding yeast, where it was first used to assess the association of differentially acetylated histones with the silent mating-type loci (Dedon et al., 1991; Braunstein et al., 1993). This method involved fractionation of cell extracts to isolate a chromatin fraction before immunoprecipitation. A closely related method was used to assess the composition of the budding yeast centromere (Meluh and Koshland, 1997). The protocol presented here (see Basic Protocol; Fig. 17.7.1) is for a simpler procedure derived by Hecht and Grunstein in which immunoprecipitations were performed with whole-cell extracts to assess the spatial distribution of SIR proteins on telomere-proximal DNA regions (Hecht et al., 1996; Strahl-Bolsinger et al., 1997). The Basic Protocol has also been applied to characterize the spatial and temporal associations of DNA replication proteins with chromatin at replication origins (Aparicio et al., 1997; Tanaka et al., 1997), the
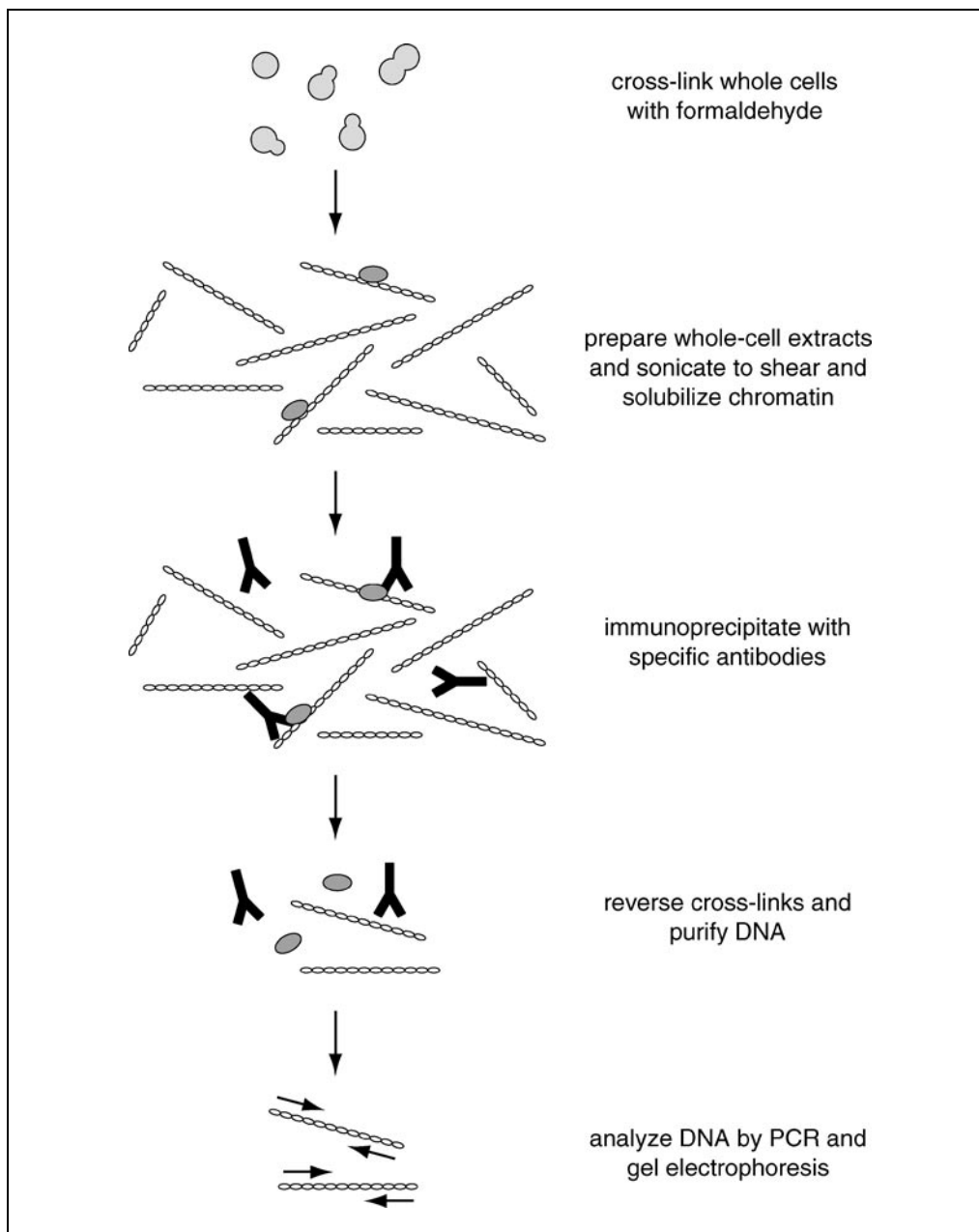
**Determining the
Association of
Proteins with
Specific Genomic
Sequences**

17.7.14

Supplement 23

Current Protocols in Cell Biology

Figure 17.7.1 Scheme for chromatin immunoprecipitation from yeast whole-cell extracts.

association of general transcription factors at promoters (Kuras and Struhl, 1999; Li et al., 1999), and the dynamics of DNA-binding proteins and chromatin-modifying activities associated with a cell-cycle and developmentally regulated promoter (Cosma et al., 1999). In each of these latter studies, protein association with relevant DNA sequences was examined using PCR amplification.

ChIP can also be used to specifically follow the genomic association of mutant proteins that are unable to support cell growth (Mencia and Struhl, 2001). This involves a strain containing both an epitope-tagged version of the mutant protein and an untagged version of the wild-type protein, which supports cell growth; ChIP is performed with an antibody against the epitope. Lastly, ChIP can be combined with microarray technology to identify the location of specific proteins on a genome-wide basis (Ren et al., 2000; Iyer et al., 2001). The immunoprecipitated DNA is PCR amplified with linkers, and the resulting material hybridized to microarrays containing the complete set of intergenic regions in *Saccharomyces cerevisiae*. Such "genome-wide location" or "ChIP-chip" is particularly powerful, because it allows one to identify novel regions of protein association, without any previous knowledge.

Macromolecular Interactions in Cells

17.7.15

Critical Parameters and Troubleshooting

Controls

There are two basic types of controls for a standard ChIP experiment. One control is mock immunoprecipitation to determine the specificity of an observed signal. This is accomplished by performing parallel immunoprecipitations of a given cross-linked chromatin sample with the antibody of interest and with an irrelevant antibody (or simply not providing any antibody). Alternatively, when antibodies against epitope-tagged proteins are employed (e.g., anti-HA, anti-myc), a similar comparison can be made with parallel immunoprecipitations (primary antibody included) of chromatin samples from strains expressing epitope-tagged or untagged versions of the protein of interest. However, in such experiments, it is often observed that the apparent immunoprecipitation efficiency for any irrelevant genomic region is about 2- to 3-fold higher in the experimental sample than in the control sample. This probably reflects nonspecific, and perhaps artifactual, association of the protein of interest with chromatin; hence it is not indicative of specific protein association with the genomic region. For proteins that generally and strongly associate with the entire genome (e.g., histones), the apparent immunoprecipitation efficiency in the experimental sample will be much higher than the control sample.

The second control, which is usually more meaningful, is to examine a given pair of input and immunoprecipitated samples for multiple genomic regions. Control genomic regions (i.e., those not interacting with the protein) should all give the same background level of apparent immunoprecipitation efficiency (typically ∼0.025% to 0.05%). Fragments bound by the protein of interest will give higher apparent immunoprecipitation efficiencies, and the difference can be expressed as fold enrichment over the background level. In the best cases, enrichments can be >100-fold, but even a factor of two can be meaningful if the experiment is repeated enough times and the experimentally determined error is sufficiently low. The advantage of this approach is that the identical samples are used to directly determine relative protein association to different genomic regions. Furthermore, differences in fold enrichments for different genomic regions represent relative quantitative measurements of protein association in vivo.

Additional controls may be used depending on the specific application. For example, where binding to a putative binding site is being tested, a mutation in the binding site is a critical control (especially if such a mutation previously was shown to eliminate binding in vitro). In a related manner, it might be useful to determine the protein association in mutant strains or under particular environmental conditions that are suspected to be important for the protein of interest.

Cross-linking

The extent of formaldehyde cross-linking is an important variable that in principle may be modified by changing the duration of cross-linking, the concentration of formaldehyde, or the temperature at which the cross-linking is performed. The use of 1% (final concentration) formaldehyde for 15 min at temperatures ranging from 12°C to 37°C usually works well; however, at temperatures above 30°C, background sometimes increases. Therefore, when fixation at a higher temperature is required, reducing the duration of cross-linking or the formaldehyde concentration may be helpful. Excessive cross-linking can interfere with cell breakage by bead beating and effective fragmentation and solubilization of the DNA by sonication (see below). For some applications where protein cross-linking is particularly efficient (e.g., histones), it might be useful to decrease the cross-linking time or formaldehyde concentration. In particular, histone tails have a number of lysine residues that are likely to be modified by formaldehyde, and such modified lysines may interfere with the binding of antibodies against specific peptides corresponding to modified histones (e.g., by acetylation, phosphorylation, methylation).

Cell lysis

Although complete lysis of all cells is not absolutely necessary (and may be difficult to achieve), it is very important that lysis be as efficient as possible. Efficient lysis is important to obtain a reproducible degree of cell breakage among a group of samples to reliably compare results. Significant differences in cell lysis efficiency will result in immunoprecipitations with different ratios of antibody to chromatin, which will possibly alter immunoprecipitation efficiency. Cell breakage by a mini bead beater is generally more efficient than breakage by a multi-vortexing apparatus, although both methods work. In both cases, it is important to use flat-bottomed 2-ml microcentrifuge tubes. When using the mini bead beater, the sample and beads should nearly fill the tube, whereas for vortexing it is important

Determining the
Association of
Proteins with
Specific Genomic
Sequences

17.7.16

to maintain a small volume. The extent of cell breakage may be monitored microscopically by comparing the number of intact cells (determined by counting on a hemacytometer) in small, diluted aliquots of the sample taken before and after vortexing. In addition, the size of the remaining pellet (unbroken cells and debris) obtained in the first centrifugation following sonication (see Basic Protocol, step 14) is a good general indicator of the extent of lysis. The size of this pellet should be routinely checked (by rapid visual inspection) to compare the extent of lysis among samples.

The final yield of genomic DNA in the extract is also an important indicator of the extent of cell breakage, although the DNA yield is also dependent on the solubilization of chromatin by sonication (see below). Poor or variable cell breakage may result from excessive cross-linking that toughens the yeast cell wall and other structures. The procedure for lysis of *Saccharomyces cerevisiae* is appropriate for other yeast species. However, modified procedures are necessary for breaking mammalian cells.

Sonication

Shearing DNA to a small size (∼500 bp average) by sonication is the critical factor in achieving resolution between a DNA sequence where a particular protein is bound and a nearby (*cis*-)DNA sequence that does not bind that protein. In addition, fragmentation of the chromatin is essential for its solubilization from the ruptured cells. As indicated above, the ability to fragment and solubilize the chromatin depends on the extent of chromatin cross-linking. In general, more cross-linking results in larger fragment size and lower solubility, resulting in lower yield (Orlando et al., 1997). Because of the importance of this variable, the shear size of the DNA should be assessed to confirm that the desired degree of fragmentation has been achieved, and it should be reassessed if fixation conditions are altered. The shear size is determined by electrophoresing DNA from step 16 of the Basic Protocol on a 1.5% to 2.0% agarose gel and visualizing with ethidium bromide. A smear of DNA should be apparent with an average size of 500 bp and most of the DNA (>90%) should be in the size range of 100 to 1000 bp.

As an alternative to sonication, DNA fragment size can be reduced by treatment of the cross-linked chromatin with micrococcal nuclease. Micrococcal nuclease preferentially cleaves DNA located in the linker regions between nucleosomes. By varying the concentration of micrococcal nuclease, it is possible to generate samples in which average DNA size varies. The minimal useful size is about 150 bp, which corresponds to a mononucleosome. However, cleavage to mononucleosome-sized fragments may also result in a preferential loss of certain genomic regions due to the sequence-specificity of micrococcal nuclease.

Immunoprecipitation

The success of this procedure relies on the use of an antibody that will specifically and tightly bind its target protein in the buffer and wash conditions used. In addition, antibody should be present in excess with respect to its target protein so that differences in the amounts of the protein-DNA complexes of interest will be accurately measured. Perform preliminary experiments to confirm avid immunoprecipitation and determine an approximate amount of antibody to use. Chromatin extracts should be prepared without prior cross-linking of the cells and subjected to immunoprecipitation with varying concentrations of antibody (20 μg/ml antibody may be a good starting point). The amount of the protein of interest in the extracts before and after immunoprecipitation should be analyzed by immunoblotting (*UNIT 6.2*) to determine the lowest antibody concentration that depletes >90% of the protein of interest from the extract. This antibody concentration is a good starting point for the full protocol and may later be modified to maximize the signal-to-noise ratio (see Anticipated Results). With cross-linked chromatin, immunodepletion of the target protein is less efficient (∼50%), presumably due to masking or modification of the epitopes, and a significant amount of the protein remains refractory to immunoprecipitation even with higher antibody concentrations. Thus, the ideal antibody concentration is ultimately determined empirically to maximize the yield of specific coprecipitated DNA while minimizing precipitation of nonspecific DNA.

Both monoclonal and polyclonal antibodies have been used in this procedure. The monoclonal antibodies 12CA5 (anti-HA), 17D09 (anti-HA), and 9E10 (anti-myc) have been used successfully in different laboratories. In general, triple-HA epitope tags work well (Hecht et al., 1996; Aparicio et al., 1997; Tanaka et al., 1997), and larger multi-myc epitope tags have also been successful (e.g., myc-9, myc-18; Tanaka et al., 1997). Protein G–Sepharose, Protein A–Sepharose, and

anti-mouse immunoglobulin-coupled magnetic beads have all been used to precipitate the immune complexes, although it should be noted that certain classes of mouse and rat immunoglobulins are not strongly bound by protein A (Harlow and Lane, 1988; see Table 7.2.1).

For optimal results, it is critical to minimize the background level of material that inevitably comes down during the immunoprecipitation. The procedures described here work well with a diverse set of antibodies, but it might be necessary to modify the binding and elution conditions in specific cases. Peptide elution is clearly preferred over heat elution, as it is more specific and results in lower experimental backgrounds and hence higher-fold inductions. However, peptide elution is only possible for experiments using antibodies against peptides (typically for analyzing epitope-tagged proteins, but analysis of native proteins should also be possible). In performing peptide elution, it is important to add enough peptide such that the protein-DNA complexes are efficiently eluted from the beads.

Another consideration is that the epitope of interest in the chromatin-bound protein might be inaccessible to the antibody due to associated proteins or DNA structures. In such a case, one might obtain a false-negative result. Whereas the majority of a given protein may be efficiently immunoprecipitated from the cross-linked cells, the fraction that is actually cross-linked might be undetectable. The use of polyclonal antibodies (which often recognize multiple determinants within a protein) or epitope-tagged proteins (the epitope is unlikely to have a specific interaction with other proteins or DNA sequences, particularly if the epitope does not affect the biological function as determined by genetic complementation) minimizes, but does not eliminate this concern. Because of this caveat, negative results should be interpreted cautiously and alternative methods (e.g., in vitro DNA binding or association of the protein with bulk chromatin) should be tried. This concern is particularly relevant when a protein of interest does not appear to interact with any genomic sequence. However, if a protein selectively associates with some genomic sequences, this concern is significantly reduced—i.e., it is unlikely that epitope masking will occur at some loci, but not others.

PCR strategy

The choice of primers depends on the experimental goals. If binding to a specific site is being tested, a primer pair that flanks the site and at least one control primer pair recognizing a DNA sequence not expected to bind the protein of interest are the minimal requirements (see Fig. 17.7.2). When choosing primers, it is important to remember that resolution between adjacent sequences is limited by the shear size of the DNA. For an average DNA size of 500 bp, a significant fraction of the DNA molecules will be in the 500 to 1000 bp
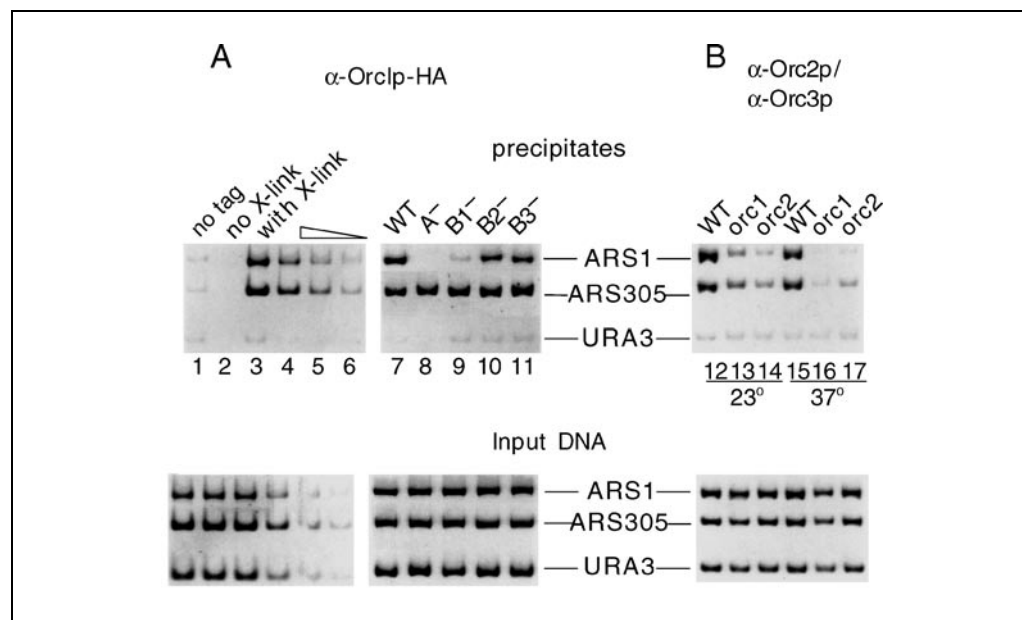

Figure 17.7.2 Anticipated results from chromatin immunoprecipitation analysis of origin recognition complex (ORC) with replication origin and nonorigin DNA sequences.

Determining the
Association of
Proteins with
Specific Genomic
Sequences

17.7.18

range, and hence DNA sequences 1000 bp distal from the actual protein binding site may be coprecipitated. Therefore, primer pairs used as controls should amplify a region of DNA that is far enough away from the expected binding site (e.g., >1 kbp) that coprecipitation of adjacent DNA is not detected. A good strategy is to design multiple sets of primers at increasing distances from a suspected binding site. Such a strategy has also been used to probe the "spreading" and "movement" of proteins on chromatin (Hecht et al., 1996; Aparicio et al., 1997; Strahl-Bolsinger et al., 1997).

Success in obtaining high-quality quantitative data is critically dependent on good primer design! In general, primers should be 20 to 30 bases long with a T_m of 55° to 60°C. Most primers require no purification or special treatment prior to PCR. Amplification products should be 75 to 300 bp. Longer PCR products should be avoided, because the amplification efficiency is substantially lower, and DNA fragments that do not bind to both primers will not be amplified (this can be a significant problem since the size of DNA fragments in the samples averages ∼500 bp and ranges between 100 to 1000 bp). A final primer concentration of 1 μM works well for most primers, but in some instances, improved product specificity may be obtained by lowering the final primer concentration 5 to 10 fold.

The design of good primers is greatly facilitated by commercially available software packages such as Oligo 6.6 or Primer Express 1.5. These packages allow for extensive customization of many different parameters, including T_m, oligonucleotide length, GC content, and more. While the success of each individual primer pair in the specific amplification of its target sequence is dependent on many variables, special care must be taken to minimize primer-dimers and hairpins. Finally, it is a good idea to check primers for hybridization to other genomic sequences through the use of a web-based program such as BLAST.

Newly obtained primer pairs must be tested for amplification specificity and performance under the conditions that will be used in quantitative PCR. Primer pairs that are suitable for reactions performed by the Basic Protocol might not be suitable for real-time PCR reactions using SYBR Green, because SYBR Green can inhibit *Taq* polymerase. It is particularly informative to analyze input DNA amplification by the primers in question on high-percentage agarose or polyacrylamide gels after completion of the PCR. The presence of multiple product bands indicates poor speci-ficity and will invariably lead to unreliable results.

For the Basic Protocol, the best test for quality of a given primer pair is to carry out a standard curve using different dilutions of DNA. For a high-quality primer pair, the amount of PCR product should be directly proportional to the amount of DNA, with an error of less than ±20%. The number of PCR cycles is determined empirically. Usually, 25 to 28 cycles is appropriate. More than 28 cycles can result in detection of nonspecifically precipitated sequences and/or lead to variable results due to inactivation of *Taq* polymerase. Multiple primer pairs can be used in combination if the PCR products are separable by gel electrophoresis (as many as five have been used), but some combinations interfere with efficient amplification of one or more products. It is essential to test primer pairs singly and in combination, with titrations of template DNA, to determine if this is a problem. The advantage of using multiple primer pairs is that individual reactions can generate data for multiple genomic regions in an internally controlled manner. In addition, the Basic Protocol can be used to simultaneously analyze two alleles of a given locus in an internally controlled manner, provided the individual alleles result in different-sized PCR products.

When quantitative PCR will be performed in real time using SYBR Green (see Alternate Protocol 2), high-quality primer pairs should result in ∼1.9-fold amplification/cycle. Such amplification efficiency can be determined from quantitative analysis of raw fluorescence data for each cycle. Amplification efficiencies <1.8 are likely to cause problems, particularly if detection of the PCR product requires 30 cycles or more. Specificity information may also be obtained by running dissociation curves on reactions following the conclusion of the PCR run.

Quantitation

For the Basic Protocol, PCR products are analyzed by gel electrophoresis and detected by staining with ethidium bromide or SYBR Green or by radioactivity (typically by including a small amount of [^{32}P]dATP in the reactions). DNA staining has the advantage of not requiring radioactivity, but the linear detection range is relatively limited. In this regard, SYBR Green is more sensitive than ethidium bromide and is preferred. Radioactive detection is more sensitive and has a larger linear range than detection by DNA staining. Whatever detection method is employed, it is

Macromolecular Interactions in Cells

17.7.19

essential to use a high-quality imager to accurately quantitate the amounts of PCR product in the various reactions.

When quantitative PCR is performed in real time using SYBR Green (Alternate Protocol 2), the linear range is directly visualized and the quality of the reactions can be directly assessed. For reactions involving a given primer, the curves should be superimposable with respect to shape, and they should differ only in the number of cycles needed to reach the threshold (C_T). Amplification efficiencies should be ∼1.9-fold/cycle. If the curves have different shapes and/or amplification efficiencies are <1.8, the reactions are not equivalent and accurate quantitation is impossible.

Data interpretation

In most experiments, it is presumed that the protein of interest associates specifically with certain genomic regions and associates only nonspecifically with other genomic regions. In general, it is very difficult to distinguish true nonspecific association from experimental background of the cross-linking procedure. In this regard, immunoprecipitations with the antibody of interest generally give 2- to 3-fold higher immunoprecipitation efficiencies than immunoprecipitations with control (or no) antibodies, but it is unclear whether this effect is physiologically meaningful or an experimental artifact.

For this reason, the best way to interpret the data for most experiments is to compare the immunoprecipitation efficiencies for different genomic regions from the same INPUT and IP samples. The immunoprecipitation efficiency is calculated by the amount of PCR product in the IP sample divided by the amount of PCR product in the INPUT sample. A typical background level for DNA fragments that do not associate with the protein of interest is 0.025% to 0.05%. However, background levels can vary, depending on the antibody used and the elution method. In general, monoclonal antibodies give lower background signals than polyclonal antibodies. Peptide elution is preferred over heat elution for the same reason, although this can only be employed for ChIP experiments involving antibodies against defined peptide epitopes. By definition, the background level should be the same for all DNA fragments that do not specifically associate with the protein of interest. In many cases, the choice of suitable negative control regions is based on expectation from other lines of evidence (e.g., the middle of protein-coding regions are unlikely to bind general transcrip-

tional initiation factors). In cases where there is no previous knowledge, the background level can only be based on multiple regions having similar immunoprecipitation efficiencies that are roughly at the level of a typical background immunoprecipitation efficiency. In this regard, it is particularly useful to use proteins tagged with a standard epitope (e.g., HA or myc), as there is considerable information on background levels in such cases.

DNA fragments that display immunoprecipitation efficiencies significantly above the background are indicative of protein association to those genomic regions in vivo. Moreover, for a given pair of INPUT and IP samples, the fold enrichment of a given genomic region over the background is directly related to the level of protein association in vivo. It is useful to define "relative protein occupancies" for different regions by subtracting the background from the observed immunoprecipitation efficiencies. For example, if the background level is arbitrarily defined as 1 occupancy unit, fragment A that shows 6-fold enrichment over background will have 5 occupancy units, whereas fragment B that shows 21-fold enrichment over background will have 20 occupancy units. Thus, one can conclude that the protein association with fragment B is 4-fold greater than with fragment A. Without further considerations (see below), this conclusion is only relevant for the particular pair of INPUT and IP samples because absolute immunoprecipitation efficiencies and fold enrichments can vary among repeated trials of the same experiment.

Absolute immunoprecipitation efficiencies and fold enrichments depend on multiple factors. First, the number and physical location of amino acid and nucleotide residues within the interacting protein surfaces that react with formaldehyde vary considerably among protein-protein and protein-DNA interactions. Second, proteins directly interacting with DNA can be cross-linked by a single event, whereas proteins that indirectly associate with DNA need multiple cross-linking events. In this regard, proteins directly binding DNA (e.g., specific DNA binding proteins and general transcription factors) typically give higher fold enrichments than other proteins (e.g., components of chromatin-modifying complexes). Third, some proteins might stably associate with genomic DNA sequences (maximally 100% occupancy), whereas association of other proteins might be transient. Fourth, the absolute immunoprecipitation efficiency depends on the quality of the specific

antibody-antigen interaction as well as the antibody concentrations, and the fold enrichment depends on both the absolute immunoprecipitation efficiency and on the background. Thus, absolute immunoprecipitation efficiencies and fold enrichments cannot be used to compare binding characteristics of different proteins. Furthermore, ChIP experiments do not provide absolute measurements of protein occupancy on specific genomic regions or relative stoichiometry of factors on a given sequence. Despite these limitations, ChIP experiments do provide direct quantitative information on the relative levels of protein association on different genomic sequences.

As mentioned above, absolute immunoprecipitation efficiencies and fold enrichments can vary among repeated trials of the same experiment due to potential differences in overall immunoprecipitation efficiency and experimental background. To account for these experimental variations, it is useful to arbitrarily define occupancy units for a specific protein-DNA association. In the example above, one might arbitrarily define fragment A to have 4 occupancy units. In independent experiments, association with fragment A will always be defined as having 4 occupancy units, and association with other fragments will be defined relative to that of fragment A in the same pair of INPUT and IP samples. In this manner, it is possible to accurately determine the relative level of protein association to multiple genomic regions even though the absolute immunoprecipitation efficiencies and fold enrichments might vary in different repeats of the same experiment. However, in actual practice, differences in absolute immunoprecipitation efficiencies and fold enrichments should not show significant sample-to-sample variation.

A related issue occurs when comparing the level of protein association under different physiological conditions (e.g., different growth conditions or strains). The ideal way to handle this situation is to analyze a "positive control" region that is predicted to be unaffected by the growth condition or genetic constitution. For example, in analyzing association of general transcription factors under a particular environmental condition, it would be useful to examine promoters that are regulated and promoters that are not. In some cases, such a control genomic region is not available, in which case one must rely on simple sample-to-sample reproducibility from independent trials of the same experiment.

ChIP can be used to determine the relative occupancy levels of different proteins at genomic regions (Kuras et al., 2000). Ideally, this is accomplished by performing parallel immunoprecipitations with different antibodies on the same cross-linked chromatin sample. Alternatively, immunoprecipitations can be performed on different samples (this often occurs when using epitope-tagged strains or when multiple proteins are examined). To determine occupancy ratios for two different factors (e.g., X and Y), occupancy units for X and Y are calculated independently as described above. The X:Y occupancy ratios are then calculated for all genomic regions examined. The resulting occupancy ratios are defined in arbitrary units, but the relative occupancy ratios for the different genomic regions are valid. To account for potential sample-to-sample variations among repeats of the same experiment, a given X:Y occupancy ratio should be defined for a specific genomic region and ratios at all other genomic regions calculated in relative terms. Using this rationale, it has been shown that the relative associations of TBP and the general transcription factors TFIIA and TFIIB are essentially constant at all promoters, whereas the TAF:TBP occupancy ratios vary considerably (Kuras et al., 2000). Importantly, however, occupancy ratios determined from such experiments cannot address whether two proteins co-occupy a given genomic region or mutually compete for the same genomic region.

For some experiments, particularly those involving histone modifications, it is inappropriate to analyze the data in terms of occupancy units and specific versus nonspecific binding sites. Histones associate with essentially all genomic regions, and the level of a particular chromatin modification typically occurs in a continuum. Thus, it is very difficult to determine whether a given region is devoid of a particular modification, although information in this regard can be obtained in control immunoprecipitations using an irrelevant antibody. For these reasons, quantitative analysis of the relative level of a given histone modification is best presented using simple immunoprecipitation efficiencies. Again, to account for sample-to-sample variations, a specific genomic region should be given an arbitrarily defined value, which is used to determine the relative levels of all other genomic regions.

Anticipated Results

Figure 17.7.2 shows the results of chromatin immunoprecipitation of protein subunits of the origin recognition complex (ORC)

and relevant controls (Aparicio et al., 1997). In panel A, immunoprecipitation of Orc1p-HA was shown to specifically coprecipitate the replication origin sequences *ARS1* and *ARS305* but not the nonorigin DNA sequence *URA3*. Enrichment of *ARS1* and *ARS305* (~0.4% precipitated relative to total) compared to *URA3* (~0.01% precipitated) was ~40-fold and depended on formaldehyde cross-linking and on the epitope-tagged Orc1p (lanes 1 to 3). Mutation of DNA sequences (A and B1) in the *ARS1* replication origin (required in vivo for origin activity and in vitro for binding of ORC to origin DNA) greatly reduced or eliminated association of Orc1p-HA with *ARS1*. The continued association with the wild-type *ARS305* origin served as an additional control (lanes 7 to 9). Analysis of the totals demonstrated that the origin and nonorigin DNA sequences were equally represented in the whole-cell extract (lower panel, "input DNA"). In panel B, origin association of ORC was tested in strains with temperature-sensitive alleles of *ORC1* or *ORC2* demonstrating loss of ORC-origin DNA binding under nonpermissive conditions (lanes 15 to 17). Nevertheless, it should not be assumed that a mutation in a protein of interest would necessarily result in loss of its chromatin association.

Time Considerations

The Basic Protocol may be completed in a 2- or 3-day period. On the first day, cells are fixed with formaldehyde and harvested (1 hr). For convenience, or if preparation of the cells for cross-linking will require an extended period of time beforehand (e.g., induction of expression, cell cycle synchronization), the cells may be frozen and stored at −80°C as described (see Basic Protocol, step 6). Preparation of chromatin extracts (2 hr) and immunoprecipitations (primary antibody incubation, ~2 hr; incubation with secondary-coupled beads, ~1 hr; washing and elution, 2 hr) may be carried out in 1 day, after which the samples are placed at 65°C overnight to reverse the cross-links. If necessary, immunoprecipitation with the primary antibody or bead-coupled secondary antibody may be extended overnight; however, it is most efficient to perform the ≥6 hr cross-link reversal step overnight. On the final day, the DNA is purified (<4 hr, including a 2-hr incubation with proteinase K), PCR amplified (<3 hr, including 2 hr for the PCR program), and analyzed by gel electrophoresis (<2 hr, including 1 hr of gel running time).

For Alternate Protocol 2, setup time is <30 min, thermal cycling takes ~2 hr, and data analysis takes <30 min.

Literature Cited

Aparicio, O.M., Weinstein, D.M., and Bell, S.P. 1997. Components and dynamics of DNA replication complexes in *S. cerevisiae*: Redistribution of MCM proteins and Cdc45p during S phase. *Cell* 91:59-69.

Braunstein, M., Rose, A.B., Holmes, S.G., Allis, C.D., and Broach, J.R. 1993. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes Dev.* 7:592-604.

Cosma, M.P., Tanaka, T., and Nasmyth, K. 1999. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle-and developmentally regulated promoter. *Cell* 97:299-311.

Dedon, P.C., Soults, J.A., Allis, C.D., and Gorovsky, M.A. 1991. A simplified formaldehyde fixation and immunoprecipitation technique for studying protein-DNA interactions. *Anal. Biochem.* 197:83-90.

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. U.S.A.* 81:4275-4279.

Gilmour, D.S., Rougvie, A.E., and Lis, J.T. 1991. Protein-DNA cross-linking as a means to determine the distribution of proteins on DNA in vivo. *Meth. Cell Biol.* 35:369-381.

Harlow, E. and Lane, D. 1998. Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Hecht, A., Strahl-Bolsinger, S., and Grunstein, M. 1996. Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* 383:92-96.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409:533-538.

Kuras, L. and Struhl, K. 1999. Binding of TBP to promoters *in vivo* is stimulated by activators and requires Pol II holoenzyme. *Nature* 399:609-612.

Kuras, L., Kosa, P., Mencia, M., and Struhl, K. 2000. TAF-containing and TAF-independent forms of transcriptionally active TBP *in vivo*. *Science* 288:1244-1248.

Li, X.-Y., Virbasius, A., Zhu, X., and Green, M.R. 1999. Enhancement of TBP binding by activators and general transcription factors. *Nature* 399:605-609.

Meluh, P.B. and Koshland, D. 1997. Budding yeast centromere composition and assembly as revealed by in vivo cross-linking. *Genes Dev.* 11:3401-3412.

Mencia, M. and Struhl, K. 2001. A region of TAF130 required for the TFIID complex to associate with promoters. *Mol. Cell. Biol.* 21:1145-1154.

Determining the Association of Proteins with Specific Genomic Sequences

17.7.22

Orlando, V., Strutt, H., and Paro, R. 1997. Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods* 11:205-214.

Ren, B. et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* 290:2306-2309.

Solomon, M.J. and Varshavsky, A. 1985. Formaldehyde-mediated DNA-protein crosslinking: A probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci. U.S.A.* 82:6470-6474.

Solomon, M.J., Larsen, P.L., and Varshavsky, A. 1988. Mapping protein-DNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene. *Cell* 53:937-947.

Strahl-Bolsinger, S., Hecht, A., Luo, K., and Grunstein, M. 1997. SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast. *Genes Dev.* 11:83-93.

Tanaka, T., Knapp, D., and Nasmyth, K. 1997. Loading of an MCM protein onto DNA replication origins is regulated by cdc6p and CDKs. *Cell* 90:649-660.

Key References

Hecht et al., 1996. See above.

Describes the technique from which the Basic Protocol was adapted.

Orlando et al., 1997. See above.

Describes formaldehyde cross-linking and immunoprecipitation for chromatin analysis in Drosophila, and discusses various parameters of the technique.

Solomon et al., 1988. See above.

Describes original formaldehyde cross-linking and immunoprecipitation technique for mapping protein-DNA interactions.

Solomon and Varshavsky, 1985. See above.

Characterizes formaldehyde cross-linking, cross-link reversal, and sensitivity of cross-linked protein-DNA complexes to proteases and endonucleases.

Contributed by Oscar Aparicio
University of Southern California
Los Angeles, California

Joseph V. Geisberg and Kevin Struhl
Harvard Medical School
Boston, Massachusetts

Macromolecular Interactions in Cells

17.7.23