January 2010

Analysis of DNA-binding Proteins in Yeast Saccharomyces Cerevisiae

Su-Wen Ho
Washington University in St. Louis

WASHINGTON UNIVERSITY IN SAINT LOUIS

Division of Biology and Biomedical Sciences

Genetics

Dissertation Examination Committee

Mark Johnston, Co-Chair
John Majors, Co-Chair
Barak Cohen
Sarah Elgin
Jeffrey Milbrandt
Gary Stormo

ANALYSIS OF DNA-BINDING PROTEINS IN YEAST *SACCHAROMYCES CEREVISIAE*

by

Su-Wen Ho

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2010

Saint Louis, Missouri

ABSTRACT OF THE DISSERTATION

Analysis of DNA-binding Proteins in Yeast *Saccharomyces Cerevisiae*

by

Su-Wen Ho

Doctor of Philosophy in Biology and Biomedical Sciences (Molecular Genetics)

Washington University in St. Louis, 2010

Professor Mark Johnston, Co-chairperson

Professor John Majors, Co-chairperson

Gene expression is an elaborate and finely tuned process involving the regulated interactions of multiple proteins with promoter and enhancer elements. A variety of approaches are currently used to study these interactions *in vivo, in vitro* as well as *in silico.* With the genome sequences of many organisms now readily available, a plethora of DNA functional elements have been predicted, but the process of identifying the proteins that bind to them *in vivo* remains a bottleneck.

I developed two high-throughput assays to address this issue. The first is a modification of the yeast "one-hybrid" assay. The second is probing protein microarrays with DNA sequence elements. Using these methods, I identified two proteins, Sef1 and Yjl103c, that bind to the same DNA sequence element.

Sef1 and Yjl103c are little-characterized members of the zinc cluster family of transcription factors of *S. cerevisiae*. Characterization of their mechanism of action as well as identification of some of their target genes leads to the conclusion that they play a pivotal role in the transcriptional regulation of utilization of nonfermentable carbon sources by budding yeast.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health Grant GM063803.

I thank the thesis committee members, Barak Cohen, Sarah Elgin, John Majors, Jeffrey Milbrandt and Gary Stormo for their support, advice and direction in this dissertation. Mark Johnston, as principal investigator, has influenced, shaped and molded both the research and the researcher through the years. His invaluable support, insights and patience are greatly appreciated.

This research could not be done without the aid of the members of the Mark Johnston, Michael Snyder and Barak Cohen laboratories. James Dover and Linda Riles in particular provided much needed assistance with experiments in the Johnston lab. Special thanks goes to Michael Snyder and Ghil Jonas at Yale University for the collaboration on protein microarrays that is integral to this research. I thank Christine Chen, Justin Gerke, Jay Gertz, Josh Witten, and Xiaowei Zhu for help with data analysis and Jon Armstrong and Ken Nelson for help in printing the protein microarrays.

We thank the Alvin J. Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital in St. Louis, Mo., for the use of the Proteomics Core, which provided MALDI-TOF data on Sef-1 and Gsm1. The Siteman Cancer Center is supported in part by an NCI Cancer Center Support Grant #P30 CA91842.

The unwavering support of family and friends, notably my husband, Jason Nowak, Jeanne Jones and Sarah Ringkor was pivotal in the final completion of this dissertation, for which I thank them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Introduction

Perspective

The Central Dogma (35) outlined in broad strokes two central concepts: sequential information transfer, and the use of a defined alphabet: four standard symbols for the components of nucleic acid. Minimal, but it was abundantly clear by that time that theses rules were sufficient to encode the probably universal set of twenty amino acids used throughout nature to create the plethora of proteins, with their diverse three-dimensional structures and myriad of functions that collectively make up a living cell (34). This paradigm created the framework for molecular biology that we still use today.

DNA-binding proteins play an integral role in the initial stages of this framework, being both responsible for replicating the genome, as well as regulating control of RNA synthesis. One of the largest and most diverse classes of DNA-binding proteins is that of transcription factors. It is the interplay between these transcription factors and their regulatory DNA sequences, each made up of a different combination and permutation of four nucleic acid bases that determines how transcription is regulated in both prokaryotes and eukaryotes.

The study of this interplay has provided a virtual explosion of information in the past two decades, and it would be impossible to cover all developments. The 2006 Nobel Prize in Chemistry brought into focus the molecular machinery involved in transcription (111). We now have the ability to visualize the orchestration of the central dogma in living cells by using fluorescence recovery after photobleaching

(FRAP) (211), thus confirming that binding of transcription factors to their DNA regulatory sequences is in rapid flux (145), and that active factors remain longer on their target sites than inactive transcription factors (260). Similar FRAP studies with fluorescently tagged subunits of RNA polymerase I expound on the role of the transcriptional machinery, showing that different subunits arrive at the bleached site at different times (233).

How transcription factors communicate information to the transcriptional machinery once bound to DNA is another rapidly evolving field. Several transcription factors, for example, Gcn4, Gal4 and Hap1 in yeast, contain separate domains for binding to DNA and activating transcription (22, 84, 99, 131); the acidic nature of these activation domains appears sufficient to cause activation (132). Other transcription factors function by recruiting other co-repressors or co-activators (167, 251). It has become clear that transcriptional regulation involves large complexes of many different proteins, which not only directly recruit components of the transcription machinery but also affect the DNA folding. Such proteins, including various chromatin-modifying enzymes, alter among other processes, nucleosome positioning and histone modifications and are potentially involved in changing the overall structure of the chromatin and/or the position of chromatin in the nucleus. Many histone acetyltransferase (HAT) complexes were first identified as co-activators (24), and the first identified histone deacetylase (HDAC), Rpd3 was identified as a co-repressor (219, 234), illustrating the fact that chromatin organization is centrally important to both gene activation and repression. Nucleosome positioning is observed at sites of activation or repression by

transcription factors (56, 124), potentially limiting accessibility of promoter elements, such as the TATA box, to *trans*-acting factors (90), or optimizing binding of the transcriptional machinery by changing the conformation of local DNA (237).

The biochemical events involving RNA polymerase II and transcriptional initiation, and the mechanism of action of individual transcription factors are two important issues that we have touched on briefly, choosing rather to focus on the initial step of how DNA-binding protein interact with their DNA sequence motifs, the ground-breaking methods used to determine this interaction for an individual protein, and the current trend of high throughput assays and computation methods to determine all DNA-binding interactions within a single organism.

Families of DNA binding proteins

DNA-protein interactions depend much upon the secondary structures that provide a surface complimentary to the structure of double-helical DNA as well as the contacts between the bases and the DNA backbone. The theory that thermodynamic interactions confer much of the stability and specificity of binding was first initiated in 1976 by the observation that two or more hydrogen bonds were necessary and sufficient for the effective discrimination between DNA bases by certain amino acids, namely that arginine at the appropriate location would recognize guanine, and that asparagine and glutamine would recognize adenine (203). While individual amino acids do confer specificity on the single base level, transcription factors can be grouped into families that use related structural motifs for recognition, and thus recognize similar groups of DNA sequences (97, 169).

These families show that there are many solutions to the structural problem of designing DNA-binding interactions, as these structural motifs have proven very successful in proliferating and adopting new roles through evolution. The three-dimensional structures of canonical members of these families have been elucidated with X-ray crystallography, allowing for greater understanding.

Helix-Turn-Helix

The first crystal structures obtained were the three bacterial regulatory proteins, CRO and CI proteins of the bacteriophage lambda (164, 168), and the CAP protein of *Escherichia coli* (144). It was apparent that they shared a distinctive string of two alpha helixes separated by a sharp beta turn (213), with a highly conserved glycine in the turn, and several hydrophobic residues in both alpha helices. Each member of this family binds as a dimer, and the approximate symmetry of the DNA-binding site is reflected in the approximate symmetry of the complex, with each monomer binding a half site. The major surface of interaction is the second helix of each monomer, docked in the major groove of each half of the binding site. The interaction is stabilized by site-specific contacts between the side chains in the HTH units and the groups in the major grove, as well as an extensive network of hydrogen bonds between the protein and DNA backbone (79).

Unlike many of the other motifs, the HTH motif is not a distinct domain, but always occurs as part of a larger DNA-binding domain. DNA sequence recognition is not only dependent on the HTH motif, but also on contacts within the larger DNA-

binding domain. For example, the CI protein augments contacts by wrapping an extended region of peptide chain around the DNA (100).

Comparative genomic studies have made it apparent that the HTH domains are present in the most prevalent transcription factors of all prokaryotic genomes (110) and some eukaryotic genomes. Evolution of the HTH domains has created sub-families with different elaborations on the basic 3-helical core. These include the tetra-helical bundle, the winged-helix and the ribbon-helix-helix (5).

Homeodomain

Unlike the HTH motif, the homeodomain forms a discrete DNA-binding domain, capable of folding into a stable structure (197). Sequence and structural similarities between the HTH and homeodomain proteins were noted with a highly conserved region of 60 amino acids (the homeobox) that appeared to contain a helix-turn-helix structure (117, 206). The crystal structures of *Drosophila* Antennapedia and engrailed proteins were among the first structures of homeobox transcription factors to be solved (108, 183). As with the HTH, many of the contacts are made by the second helix in the major groove. However, as this helix is much larger than the corresponding helix in the HTH motif, different parts of each helix are closest to the DNA bases, thus suggesting that corresponding residues do not make critical contacts (108). Although an isolated homeodomain binds DNA with similar sequence specificity to the intact protein, flanking sequences that are conserved with different subfamilies may modulate binding (202). Homeodomain proteins bind as both

homodimers and heterodimers, increasing the diversity of DNA site preferences (105).

Leucine Zippers

The leucine zipper was first discovered as a conserved sequence pattern in several eukaryotic transcription factors (116). These transcription factors contain two subdomains: a dimerization domain with the hydrophobic amino acid leucine (leading to the name leucine zipper) at every seventh position, and a basic region that contacts DNA. The leucine zipper region forms two parallel α–helices in a coiled-coil arrangement (185).

Binding of the basic region is not dependent on the leucine zipper domain, since the basic region of the yeast transcription factor, Gcn4 is able to bind specifically as long as a disulfide bond is added to allow dimer formation (218). A high-resolution structure of Gcn4 indicates that the dimeric protein contains two extended α-helices that interact with DNA at two adjacent major grooves separated by about half a turn of the double helix, reminiscent of a pair of scissors (163). Binding is due to interaction of the basic residues with the phosphates in the DNA backbone, and with specific bases in the major groove. Leucine zipper proteins can form heterodimers, thus acquiring new DNA-binding specificities (76), and new regulatory functions (186).

As additional family members were identified, it was discovered that they contained other hydrophobic amino acids in the position of the heptad of leucine residues. These proteins also contained a C-terminal coiled-coil dimerization domain

and a basic DNA-binding domain. The term basic zipper (bZip) is now used to refer to this structural class of proteins (44, 236).

Helix-Loop-Helix Proteins

The helix-loop-helix (HLH) proteins appear similar to the leucine zipper proteins (157, 158), having a dimerization domain and a basic DNA-binding domain. The difference is that the dimerization region forms an α-helix, a loop and a second α-helix. As with leucine zipper proteins, HLH proteins form heterodimers, and have many different roles in differentiation and development, exemplified by MyoD, the primary signal for differentiation of muscle cells (249, 250).

Zinc Finger Proteins

A number of different proteins have regions that fold around a central Zn^{2+} ion, producing a compact domain from a relatively short length of cysteine-rich polypeptide chain. Binding of the Zn ion to cysteine and histidine residues stabilizes the domain and contributes to proper protein function and structure (115, 253). This superfamily of proteins is not limited to transcription factors, having many other physiological roles including mediating protein-protein interactions, chromatin remodeling, protein chaperoning, lipid binding and zinc sensing (115). The three common sub-classes that bind DNA are the C_2H_2 zinc-finger domain (253), the C4 zinc-finger (109), and the C6 zinc finger or zinc cluster proteins (135).

The **C_2H_2 zinc fingers** are one of the most common motifs in eukaryotes, being used not only for protein-DNA interactions, but also for protein-protein

interactions and protein-RNA interactions (134). They were first identified in the *Xenopus* transcription factor TFIIA that contains nine tandem repeats of the approximately 30 amino acid motif (151). The name "zinc finger" was coined because a two dimensional diagram of the structure resembles a finger, consisting of one helix and a pair of anti-parallel β strands (253). Each C_2H_2 finger has the consensus sequence Tyr/Phe-X-Cys-X_{2-5}Cys-X_3-Phe/Tyr-X_5-φ-X_2-His-X_{3-4}-His where φ is a hydrophobic residue. A $zinc^{2+}$ ion binds between the two cysteine and two histidine residues, allowing the polypeptide to fold into a compact domain that can insert its α helix into the major groove of DNA (53). There are very few fully conserved residues in the zinc fingers because the intrastrand "cross-linking" by the zinc ion provides most of the structural stability (149). A number of studies have tried to determine the principles of DNA recognition of these zinc fingers, both experimentally (159, 160), and computationally (258). The variety of DNA sequences that the proteins bind to is determined by the combination of three or more repeating C_2H_2 fingers that interact with successive groups of base pairs as the protein wraps around the DNA double helix. A subclass within this group is the FOG (friends of GATA) proteins that contain the C_2H_2 fingers as well as a C_2HC consensus sequence (224). **C_2H_2 zinc fingers** usually bind as monomers (115).

In contrast, the **C_4 zinc finger proteins** generally contain only one finger unit binding DNA as homodimers or heterodimers. The first members of this class were identified as mammalian hormone receptors (247, 248). The GATA transcription factors are a key example of this class of transcription factors, and bind to the GATA motif through two zinc finger domains (227). The consensus sequence Cys-X_2-Cys-

X_{13}-Cys-X_{14-15}-Cys-X_5-Cys-X_9-Cys-X_2-Cys contains two groups of four cysteines, each group binding a Zn^{2+} ion. Like the HTH homodimers, C_4 zinc finger homodimers have two-fold rotational symmetry, and therefore recognize DNA binding sites with inverted repeats, whereas the heterodimers bind to direct repeats (115).

Unlike the C_2H_2 family of zinc finger proteins that are prevalent in eukaryotes ranging from yeast to humans, members of the **C6 zinc cluster protein** family are exclusively fungal and have the conserved motif Cys-X_2-Cys-X_6-Cys-X_{5-12}-Cys-X_2-Cys-X_{6-8}-Cys (135). The name "zinc cluster" stems from the binding of the six cysteines to two zinc atoms to form a single zinc finger unit with a cloverleaf-shaped structure (170). These proteins are unique in that they may contain a single zinc finger that binds two zinc atoms. The first and fourth cysteines act as bridging ligands by ligating both metal ions while the remaining four cysteine residues act as terminal ligands (57, 171). This motif can be considered as two Cys-X_2-Cys-X_6-Cys repeat units separated by a spacer of six residues. Each unit forms a short α-helical structure separated by a loop containing a conserved proline residue that confers flexibility (10).

Zinc cluster proteins can interact with DNA as monomers, homodimers, and heterodimers (1, 221). The most well-known and well-characterized member of this family is the *S. cerevisiae* transcription factor Gal4 (99, 128). The zinc -binding cluster lies in the DNA major groove and contacts three base pairs (10). This trinucleotide sequence is often a CGG triplet and zinc cluster proteins recognize highly related elements. The spacing and orientation of these CGG triplets determines

which family member binds to the sequence (75, 80, 126). For example, the Hap1 DNA-binding domain binds CGG in a direct repeat, Ppr1 and Put3 bind to an inverted repeat and Leu3 binds to an everted repeat. The critical nature of spacing is illustrated by comparing the Gal4 binding site ($CGGN_{11}CCG$) and the Put3 binding site ($CGGN_{10}CCG$) (7, 99).

The similarity of binding sites of various family members not only reflects the high homology between members but also suggests that other factors must influence DNA targeting. Indeed, at least two known zinc cluster proteins, *S. cerevisiae* Dal81 and *Aspergillus nidulans* TamA are fully functional even when their zinc clusters are deleted (23, 40), and three other members (RSC3, RSC30 and Cep3) do not bind to DNA directly (4, 121). Moreover, swapping their zinc fingers does not appear to affect DNA targeting (187).

In addition to the zinc fingers, the DNA-binding domain is separated into two other regions: the linker region and the dimerization domain. The linker region is located C-terminal to the zinc cluster motif. It can take on many different forms and appears to contribute to binding specificity (136). For example, the linker region of Ppr1 is an antiparallel β sheet (140) while the Gal4 linker region extends along one DNA strand, contacting the phosphodiester backbone (139). However, mutations in either linker region affect DNA binding and protein function of the respective proteins (98).

The dimerization domain is made up of heptad repeats, similar to those found in leucine zippers (198), that form a highly conserved coiled-coiled structure responsible for dimerization and protein-protein interactions.

These major families make up the bulk of transcription factors. To date, more than 100 different DNA binding domains have been found (112). These domains have been used computationally to predict transcription factors in a genome of interest. For example, *S. cerevisiae* encodes ~200 predicted transcription factors (112, 122), *C. elegans* contains 934 predicted transcription factors (188) and more complex eukaryotes such as humans may use up to 10% of their coding genome to code for transcription factors (123).

The DNA that transcription factors bind to

A gene promoter is the regulatory sequence directly upstream of the transcription start site. In *S. cerevisiae*, the identification of promoters is relatively straightforward as the genome is compact (67), with few introns and short intergenic regions, most under 1000 base pairs. In more complex genomes, longer intergenic regions with many repeat sequences, ill-defined transcription start sites and multiple alternative promoters make promoter identification more difficult. Several experimental approaches including full-length cDNA sequencing (89) and chromatin-immunoprecipiation (ChIP) with anti-TFIID and anti-RNA polymerase antibodies (107) have provided some definition, but more sensitive methods are needed.

Located within the promoters are individual *cis*-regulatory elements that transcription factors bind to. These short elements (usually <20 base pairs) often occur in clusters and in combination with binding sites for other transcription factors. Binding sites tend to be degenerate, with the degree of degeneracy thought to reflect the type of protein-DNA interaction at each position (152). The most popular method

to profile a binding site is a position weight matrix (PWM), created by aligning identified sites and counting the frequency of each DNA base at every position of the alignment (36, 59, 215). According to this model, each base of the site contributes independently to the binding of the transcription factor (13). This assumption, while incorrect (25, 137), is not fatally inaccurate (12).

Several hundred matrices for specific transcription factors are collected in databases such as TRANSFAC (http://www,biobase.de/) (252) , UniPROBE (http://thebrain.bwh.harvard.edu/uniprobe/) (161), and JASPAR (http://jaspar.genereg.net) (238). However, reliable prediction of sites in long sequences is near impossible as many of these available binding matrices are too small and not specific enough (231). As more binding sites for specific transcription factors are identified by computational and experimental methods, the predictive success of these matrices can only improve.

Identification of binding sites by computational methods

Two general computational methods have emerged to identify transcription factor binding sites in promoters *de novo*: analysis of co-regulated genes and phylogenetic footprinting. The first looks for recurring or overrepresented sequences in promoters of genes that are similarly expressed (184, 222). Examples include Hidden Markov Models (175), Gibbs sampling (118), expectation-maximization (MEME) (9) and greedy alignment algorithms (CONSENSUS) (82). These methods differ in how binding profiles are represented and in the assumptions that they make regarding the presence and position of the binding sites in the promoters.

The second method, phylogenetic footprinting, is based on conservation of functional elements in closely related species. The filtering power of evolutionary constraints allows binding sites and other functional *cis*-regulatory elements to stand apart from background sequence conservation. This method has been used to identify putative elements in yeast (31, 32, 106), *Drosophila melanogaster* (66), fish (143) and human genomes (11, 256). Analysis tools that have been refined by including conservation include Gibbs sampling (PhyloGibbs-MP) (207), and greedy alignment algorithms (Phylocon) (245).

Despite the best efforts in predicting functional sites, the cellular environment still dictates which binding events can and cannot occur due to a myriad of environmental constraints; thus, experimental confirmation still remains the highest form of validation. Described below are various experimental techniques that can be used to identify and confirm DNA-protein interactions. Experimental methods can range from localized, site specific analysis of a single transcription factor and its binding site on a given promoter to high-throughput methods which lose sensitivity, but gain by generating broad conclusion about binding site preferences and regulation of gene expression.

Identification of binding sites by experimental methods

Protein-DNA interactions can be mapped using two conceptually different strategies. One can identify a transcription factor of interest, and use it to pinpoint the DNA that it binds to. Conversely, one can take a DNA sequence and use it to identify

the transcription factor that binds to the sequence. We refer to these methods as protein-centered and DNA-centered respectively.

Protein-centered experiments

Traditional *in vitro* methods of studying DNA-protein interactions include electrophoretic gel mobility shift assay (EMSA) (54, 58), nitrocellulose filter binding (172, 255), Southwestern blotting (19, 114), and DNA footprinting (55, 120). In these methods, DNA is typically labeled with radioactivity to aid in visualization. However, there are now also a number of non-radioactive alternatives that avoid the use of radioisotopes (30, 37).

Both **EMSA** and **filter binding assays** are powerful methods based on the principle that DNA-protein complexes migrate differently from free DNA. In the former, the DNA protein mixture is separated by gel electrophoresis on a polyacrylamide gel and visualized using labeled DNA. DNA-protein complexes migrate slower than free DNA, and thus forming a band that is "shifted". Use of antibodies to the specific protein will retard the complex further, causing a "supershift".

Using a wide range of buffer conditions, nucleic acids pass freely through nitrocellulose membranes while proteins and their bound ligands are retained. Thus, if a specific protein binds to a specific DNA sequence, passage through the filter will result in the retention of a fraction of the DNA-protein complex. The amount of DNA

retained can then be determined, allowing a binding curve to be constructed. Both techniques are suitable to qualitative, quantitative, and kinetic analyses.

In contrast to the two previous methods, DNA-protein binding is the last step in **southwestern blotting**. Proteins are first separated on a sodium dodecyl sulfate (SDS) polyacrylamide gel, then renatured in SDS-free buffer and transferred by electroblotting to an immobilizing membrane, and detected by their ability to bind labeled DNA. This combines the advantages of a high-resolution fractionation step with the rapid analysis of a large number of different DNA-binding proteins.

The association of proteins with the DNA double helix can interfere with the accessibility of the latter to nucleases and other footprinting agents. This is particularly true when using DNAse I, which is bulky and is relatively easily sterically hindered. The **DNA footprinting** method was developed to take advantage of this phenomenon. In this method, a sequence of DNA is uniquely end-labeled and partially digested in the presence or absence of a specific DNA-binding protein. The two sets of fragments are then separated side by side on a gel and the patterns compared. The region of protection will show up as a gap (or footprint) in the otherwise continuous background of digested products. Certain areas may show enhanced cleavage, indicating increased availability for digestion due to changes in DNA structure. This technique can reveal if multiple binding sites for the same protein are present on the same fragment and allow the comparison of their respective affinities. Other footprinting agents include Exonuclease III (174), diethyl pyrocarbonate and potassium permanganate (102), uranyl(162), ethylnitrosourea (138) and hydroxyl radicals (91).

The development of *in vivo* footprinting now allows the study of DNA-protein binding event within a living cell. This assay uses ligation-mediated PCR (**LMPCR**) to capture the fractured pieces of genomic DNA that flank the sites protected by the protein (46, 154).

The methods outlined above use predetermined DNA sequences to look for interaction with the transcription factor of interest. Systematic Evolution of Ligands by EXponential enrichment (**SELEX**) (18, 165) selects high affinity binding sites for the specific transcription factor from a pool of often-random DNA sequences. Briefly, SELEX involves three processes, namely: selection of ligands that bind to a target protein, separation of bound complexes from unbound DNA, and amplification of the bound sequences. Through repeated amplification and several selection cycles, the DNA sequences that bind with high affinity and specificity to the target protein are enriched. This method, first used for DNA and RNA binding proteins, has since been used for the selection of nucleic acid ligands for any kind of targets (68).

Reporter genes, such as β-galactosidase, are used in many *in vivo* assays, including a yeast-1 hybrid (reviewed below) and deletion analysis of promoters (e.g. (205)). In one such assay, the DNA-binding domain of the protein of interest is expressed in yeast cells as a fusion with a known transcriptional activation domain and the target binding site is used as an artificial activation sequence (UAS) in an engineered promoter driving expression of a reporter gene. Expression of the reporter gene is dependent upon specific, high-affinity interaction between the synthetic UAS and the DNA-binding domain of the artificial activator (204). In bacterial one-hybrid assays, a transcription factor of interest is expressed in bacteria containing a library of

random DNA elements in front of a reporter gene that allows growth under selective conditions when the transcription factor binds to the element (147, 148). The recognition sequence of the transcription factor can be derived by alignment of the DNA elements from multiple selected colonies.

High-throughput protein-centered assays

Microarrays have been used in both protein-centered and DNA-centered methods. In the protein-centered method, a purified transcription factor fused to glutathione S transferase (GST) is incubated with a double stranded DNA array (155). This method has been used to find targets for Abf1, Rap1 and Mig1, and the target sequences used to identify the consensus binding sites for each of these factors. Recently, a DNA binding survey of yeast transcription factors was done using amore than 2.3 million gapped and ungapped 8 basepair sequences to determine high – resolution profiles for 89 known and predicted yeast TFs (265). A complimentary method is **DIP-ChIP** in which naked genomic DNA is incubated with a purified transcription factor, and the resultant complexes are immunoprecipitated (127). The sequences that bind to the transcription factor are identified by microarray analysis. Although both methods are carried out *in vitro*, the binding sites obtained are in good agreement with those obtained from *in vivo* assays.

Many protein-DNA interaction mapping methods are based on chromatin immunoprecipitation (**ChIP**) (141). ChIp assays are a modification of "pull-down" assays in which target proteins are retrieved using an antibody coupled to a retrievable tag. In contrast to standard immunoprecipitation assays, ChIp assays capture *in vivo* protein-DNA interactions by crosslinking proteins to their DNA using

formaldehyde or other crosslinking agents such as UV (264). These DNA fragments can subsequently be identified and quantified using a variety of readouts including PCR, microarrays (**ChIP-chip**) (85, 190) and "next generation" DNA sequencing (122, 192) (**ChIP-PET** and **ChIP-STAGE**). For yeast ChIP-chip assays, endogenous transcription factors were replaced by hybrid proteins in which the transcription factors were fused to the same universal tag (122). Almost 200 transcription factor fusions were created, allowing query with the same antibody for each transcription factor. Target binding under standard lab conditions as well as multiple experimental conditions have been tested (78, 122, 257). ChIp-chip has also been successfully applied to map the target genes in other organisms (20, 263)).

Variations of ChIP use other methods of shearing DNA in order to analyze insoluble proteins, such as the scaffolding components of chromatin. This includes using micrococcal nuclease tethered to an antibody (**ChIC**) and **ChEC,** which uses the DNA-binding protein itself (199).

Two recent methods (**Calling Card** and **DamID**) use transcription factors fused to proteins that modify DNA in order to identify genomic sites where the transcription factors bind. In **DamID**, *E. coli* DNA adenine methyltransferase (Dam) is fused to a transcription factor and expressed in intact cells (229, 230). Upon binding of the transcription factor to DNA, surrounding adenines are methylated. DamID has been used to dissect the *Drosophila* Myc transcription factor network (166). The **calling card method** uses Sir4 fused to the transcription factor of interest (242, 243). When the fusion protein binds to a site in the genome, it recruits the Ty5 integrase and thereby directs insertion of Ty5 into the genome. Analysis of sequences

surrounding the Ty5 insert allows identification of the promoter region. This method has been used successfully to identify binding sites for Gal4, Gcn4, Pho4 and Pho2. The calling card method is one of the few that attempts to be a DNA-centered assay in its goal to identify all transcription factors that bind to a promoter of interest. To this end, each DNA-binding Sir4 fusion protein is provided with a unique bar-coded Ty5 calling card. A mixture of strains is used and all proteins that bind to a particular region of the genome can be identified by recovering the Ty5 elements deposited in the region and by reading the bar code sequences that they carry.

DNA-centered methods

The wealth of putative binding sites derived from computational data needs to be validated by experimental methods. The expression of the regulator itself may not correlate with expression of its target (e.g. if the transcription factor is regulated post-translationally), as significant correlations between known transcription factor-target pairs are infrequent (81, 182). Unless a candidate transcription factor can be identified, protein-centered DNA-protein interaction assays cannot be used. However, DNA-centered methods are much less common (48, 119, 240). To date, there are only four published cases in which a binding site was discovered computationally, and its DNA-protein interaction experimentally demonstrated (61, 83, 142, 153).

Traditional methods for the unbiased identification of sequence specific DNA binding proteins use a combination of several steps of classical chromatography followed by an affinity purification step that uses the recognition sequence as a ligand (101). This classical approach is laborious and requires monitoring the purification

process by functional assays (e.g. EMSA) and is thus impractical on a proteomic level.

Four genomic collections have been made that express yeast ORFs fused to purification tags (176), which in theory, could be used to take this method to the proteomic level by high-throughput purification of the tagged proteins. By pooling the tagged yeast strains and assaying for binding function of the purified pooled proteins to a specific DNA sequence (e.g. by EMSA), one could quickly narrow down the transcription factor responsible for the binding activity. These proteome libraries have already been used to identify other protein activities, such as RNA-modification (94). However, this method has yet to yield a specific DNA-protein interaction, and the only observed binding protein in two such experiments was a non-specific DNA binding protein, Apn1 (data not shown and E. Phizicky personal communication).

The **calling card** method described above is one of the few DNA-centered methods, and is only applicable to yeast. The **yeast one hybrid** uses yeast as a tool, but can be used to map regulatory pathways in other organisms, for example, *C. elegans* (43). Yeast one hybrid was first developed to identify proteins that can bind to multiple copies of a short DNA sequence (125, 244). The yeast one hybrid is similar to the bacterial one hybrid, except that a library of proteins is fused to an activation domain (AD) and used to query a single sequence. Under selective conditions, strains containing protein-AD fusions that bind to the sequence can be identified by PCR and sequencing. The use of the AD fusion enables identification of both activators and repressors.

A high-throughput version of the yeast one hybrid compatible with the Gateway cloning system has been established (42). The Gateway system is a recombinatorial cloning system that allows many DNA sequence fragments to be cloned simultaneously(241). This system also makes use of Gateway compatible "protein prey" resources. Mini-libraries consisting solely of predicted transcription factors can be created and screened.

I developed a modified yeast one hybrid that was used to identify two different transcription factors that bound to DNA sequences predicted by computational methods (61). This will be discussed in chapter 2.

A very recent invention is the Stable Isotope Labeling by Amino acids in Cell culture (**SILAC**) based DNA protein interaction screen (153). Briefly, proteomes are metabolically labeled with $^{2}H_{4}$-lysine to allow discrimination in peptide mass. Biotinylated DNA probes containing the binding sites are synthesized and immobilized on strepavadin magnetic beads. The nuclear extract is subjected to DNA affinity chromatography and the resultant purified proteins are identified by mass spectrometry (MS). Unlabeled nuclear extract purified with DNA affinity chromatography using an unrelated DNA sequence is used as a control to eliminate proteins that bind non-specifically. This protocol was used to identify several proteins that bound to the methylated CpG island upstream of a human gene promoter.

The other DNA-centered assay is **protein microarrays** (83) (214) (86). This is the converse of the protein-centered DNA microarray assays. We developed the first protein microarray assay to test for DNA-protein interactions using a yeast transcription factor microarray. This will be discussed in chapter 3. All known and

putative transcription factors in yeast were purified and used to create a protein microarray. The microarray was queried by a labeled DNA sequence, and proteins that bind to the sequence identified. The known binding sites for Rap1 and Abf1 were identified as well as a novel DNA-protein interaction.

The techniques that we developed with our proof of principle experiments have since been used to create an *Arabadopsis* transcription factor microarray which has been used to identify four novel transcription factors that bind to the evening element (EE) (69). Commercially available protein microarrays containing many classes of proteins have also been employed to identify protein interactions with metal-modified DNA (210).

Using the methods discussed in chapter 2 and 3, we identified 2 DNA-binding proteins, Yjl103c and Sef1 that bound to the same DNA sequence motif $CCGN_8CCG$. In chapter 4, we discuss several experiments used to determine the function of these proteins.

The interaction between transcription factors and their DNA binding sites are an integral part of gene regulatory networks and represent a key interface between proteome and genome. The burgeoning field of systems biology is filled with attempts to model the physical and regulatory interactions between transcription factors and their target genes (reviewed in (240)). We have developed two high-throughput methods that can only aid in this process.

CHAPTER 2

Linking DNA-binding proteins to their recognition sequences by a modified yeast one hybrid method

(The work on Stp2 was published in Genome Res. 2005 August; 15(8): 1145–1152)

Abstract

We have developed a modified yeast one-hybrid assay (MY1H) useful for high throughput identification of DNA-binding proteins that bind to a specific DNA sequence motif. Using a promoter with zero background expression, we vastly reduce the number of false positives that bind to a given sequence. This technique was used to identify two different protein-DNA interactions. Stp2 was identified in a screen using a sequence motif derived computationally from a study of co-expressed genes. Sef1 was identified in a screen using sequence motifs that were conserved through evolution.

Introduction

Gene expression is an elaborate and finely tuned process involving the regulated interactions of multiple proteins with promoter and enhancer elements. A variety of approaches are currently used to study these interactions, *in vivo, in vitro* as well as *in silico.* The yeast one-hybrid system (Y1H) is a frequently used genetic assay to identify protein-DNA interactions(244). This variant of the yeast two-hybrid (Y2H) system (49) is useful for isolating genes that encode proteins that bind to *cis*-acting elements and for further characterization of known protein-DNA interactions (3, 125), whereas the Y2H allows detection of protein-protein interactions(173). Both are powerful tools that can be used in high-throughput assays to aid in mapping cellular networks (92, 130, 226).

I have developed a modified Y1H system (MY1H) useful for high-throughput examination of protein-DNA interactions. In order to cut down the number of false positives, a zero-background promoter construct was designed to control expression of the HIS3 gene. Although the expression of a reporter protein is an indirect measurement of the transcriptional properties of the test DNA, it is generally proportional to transcriptional activity (2). Ideally, expression of the reporter gene would only occur when an upstream activating sequence (UAS) is inserted into the core promoter sequences and under conditions when this element is active. The pioneering work of L. Guarente and colleagues used DNA sequences upstream of the yeast *CYC1* gene and deleted regions containing all UAS activity (72-74). This promoter construct has been widely used (50, 201) as the majority of yeast genes

have TATA boxes that overlap with positioned nucelosomes and are therefore inaccessible in the absence of transcriptional activators (113). In contrast, one of the two TATA boxes of the *CYC1* promoter is constitutively accessible (29), which could explain residual reporter gene expression from the CYC1ΔUAS plasmids even in the *absence* of inserted UAS elements. We took advantage of the tight *MEL1* core promoter to create a zero-background promoter construct so that *HIS3* is expressed only if a *cis*-acting DNA element is inserted into the promoter and only when its interacting protein-Activation Domain (AD) fusion is present. *MEL1* is regulated by a single Gal4 binding site and there is no detectable Mel1 activity in a *gal4*Δ strain (181).

To make the system high-throughput, a strain carrying the promoter construct on a plasmid is mated with a library of strains carrying AD-fusion proteins, allowing for rapid identification of interacting proteins.

Using this method, we were able to identify several proteins that bound specifically to different DNA sequence elements identified by computation methods. In particular, we identified two novel interactions for Sef1 and Stp2.

Results

Design of zero-background promoter construct

In order to reduce the number of false positives, we designed a zero background promoter construct based on the *MEL1* promoter. Expression of Mel1 is tightly regulated by two *cis*-acting sequences: a repressive Mig1 binding site, and an activating Gal4 binding site (146, 181). By homologous recombination, we removed both cis-acting sequences and created a *MEL1* minimal promoter interrupted by the *TRP1* gene, flanked by SpeI and XhoI restriction sites. (Figure 2_1A) Non-recombinant plasmids can be counterselected using 5-fluroanthranilic acid (223), improving the efficacy of insertion of desired DNA sequence elements into the promoter by gap repair. HIS3 was used as the reporter gene as titration of 3-aminotriazole (3AT) can be used to eliminate false positives (38).

Design of DNA elements used for insertion into promoter construct

We inserted desired *cis*-acting DNA elements into the promoter construct by homologous recombination using gap repair with a double stranded DNA fragment created using two long oligonucleotides in a simple fill-in reaction. (Figure 2_1B). The resultant DNA fragment contained flanking 20 base pairs of homology to the MEL1 promoter for efficient homologous recombination. In addition to these regions of homology, DNA sequences containing more than one putative *cis*-acting element include an 18 base pair spacer comprising of a BamHI restriction site and a 12-mer DNA sequence shown to be absent in all known regulatory regions in the

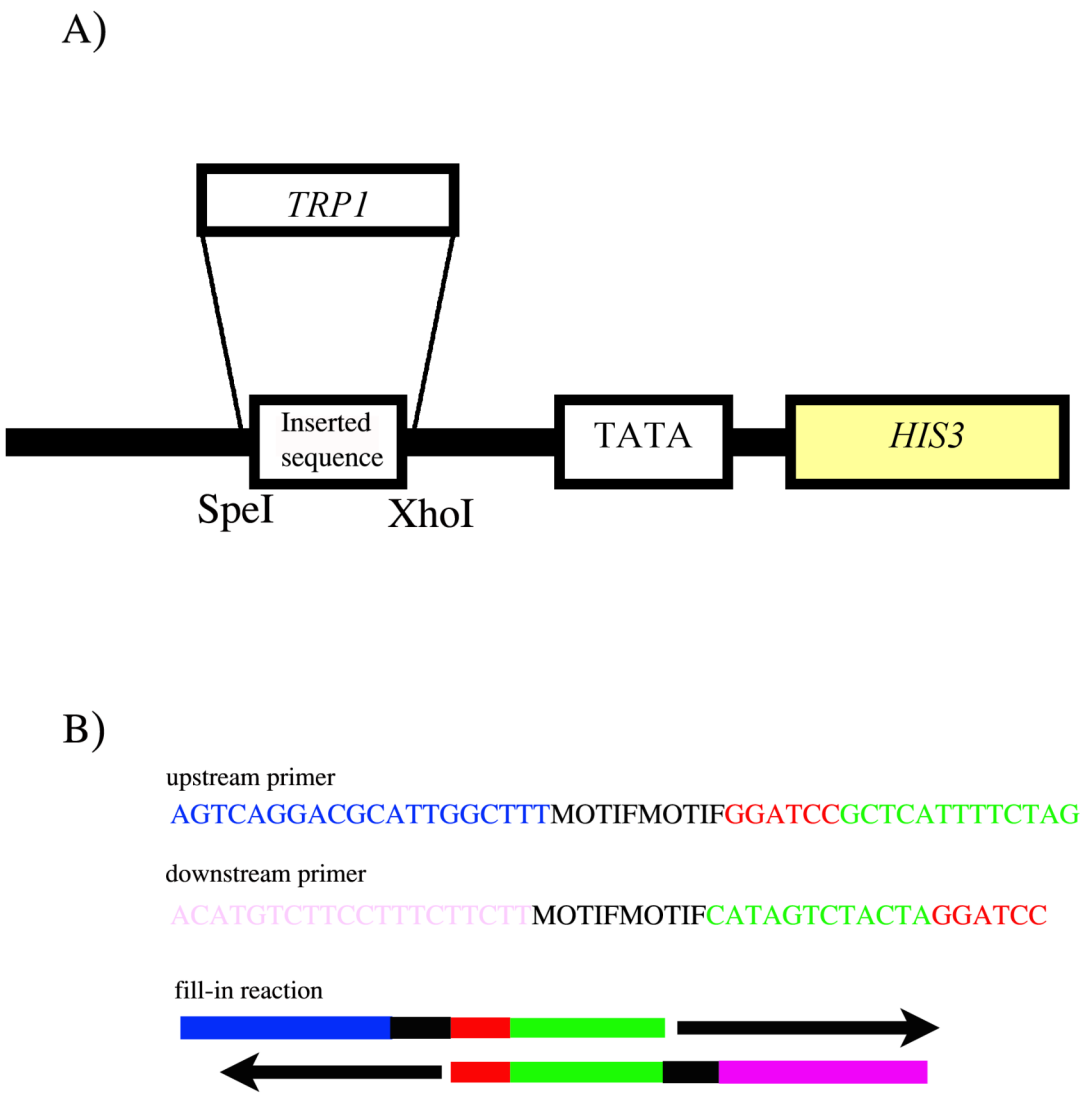


Figure 2_1: (A) Zero background promoter construct. *TRP1* is excised by cutting with SpeI and XhoI, and the desired motif sequence is inserted by gap repair. **(B)** Creation of double stranded DNA for insertion into promoter construct. The two oligonucleotides anneal together using the BamHI site (in red) and the unique sequence (in green). A simple fill-in reaction creates a double stranded fragment for insertion.

S. cerevisiae genome (John Majors, personal communication). This allows us to increase the likelihood of protein binding as well as the possibility of incorporating different DNA elements into the same promoter construct.

Transcription factor AD-fusions

We used 169 different strains expressing transcription factor fusions to test the promoter constructs (49) for the study of Stp2, and a larger group of 269 strains for the second study (Supplemental Tables 1 and 2). This was a good representation of transcription factors in *S. cerevisiae* at the time. Confirming the size of the DNA encoding the proteins ensured the fidelity of these fusion proteins. Fusion proteins that did not meet our stringent criteria were left out. These strains were pinned in 96 grid format and were crossed to a single strain carrying the reporter plasmid. The resultant diploid cells were scored for histidine prototrophy.

Proof of principle

In order to test out our system, we inserted into the promoter construct known binding sites for several yeast transcription factors, including Rgt1, Ume6, Cin5, Gcn4 and Rpn4. Insertion of the additional sequences in some cases was sufficient for slight *HIS3* expression. This background was easily removed by the use of a low dose of 3AT (5mM) in the growth media. In a few cases, activation occurred even at high doses of 3AT. This activation was not necessarily caused by the endogenous transcription factors that we were testing, as activation occurred even in the strains

lacking the corresponding transcription factor (data not shown). The strain carrying the Ume6 binding motif did not grow under any conditions with the Ume6-AD fusion. This may be due to the fact that Ume6 is a meiosis specific transcription factor, and may require meiosis-specific conditions to bind to its site.

While self-activation in many cases precludes the use of these promoter constructs in the MY1H assay, as all query strains show up positive, others, including constructs containing Cin5 binding, still show specific activation with their corresponding transcription factor-AD fusions (Figure 2_2B). Specific binding of Rpn4 to its known binding sites was also observed (Figure 2_2A).

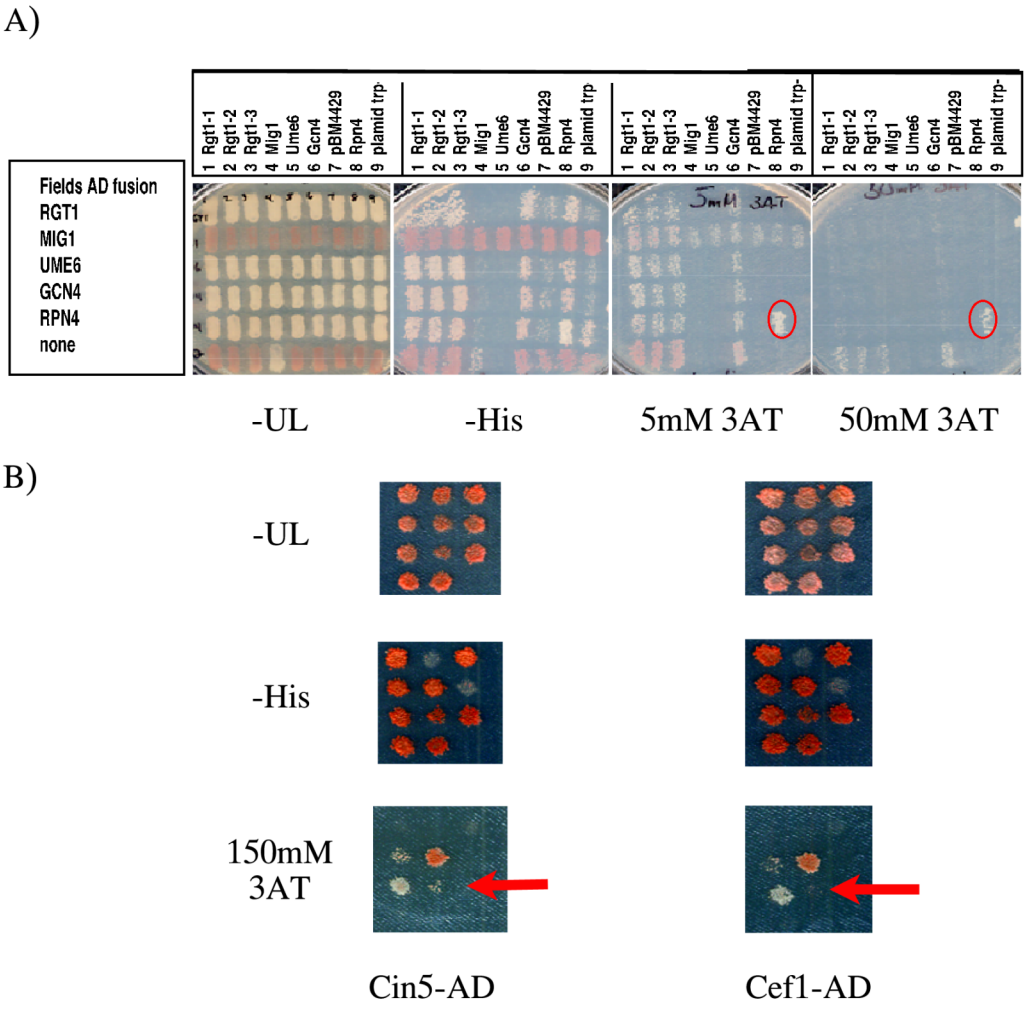


Figure 2_2: Proof of principle

(A) 9 strains carrying a different reporter plasmid were crossed with 6 strains expressing protein-AD fusions. Strains carrying Rgt1 binding motifs grow under all conditions, as does the strain carrying the Gcn4 binding motif. The strain carrying the Ume6 binding motif did not activate under any conditions. The strain carrying the Rpn4 binding motif only grows when crossed to the strain expressing the Rpn4-AD fusion (circled in red).

(B) 11 strains carrying a different reporter plasmid were crossed with a strain expressing a single protein-AD fusion: Cin5-AD on the left and Cef1-AD on the right. The arrows point to the strains containing the Cin5 binding motif, indicating that Cin5-AD fusion activates expression specifically from the Cin5 binding motif.

Discovery of Stp2 binding site by MY1H

The motif

A promising candidate sequence motif was identified using a novel algorithm that searches for short conserved sequence motifs in the genomes of related species (61). The sequence logo (200) was further refined by additional rounds of selection (Figure 2_3A). The refined motif is conserved in 19.9% of promoters when the *S. cerevisiae* promoter has a site, which is comparable to known binding sites (i.e., Ume6 binding site is conserved 20.0%). The genes whose promoters contain copies of this motif show coherent expression (EC = 0.38, $P < 10^{-6}$) (177) in cells treated with the DNA damaging agent methyl-methane sulfonate (MMS) (96). The promoters that contain this binding site also overlap significantly with those identified in ChIP experiments (122) with Sfp1 ($P = 0.00035$), Stp2 ($P = 0.00011$), and Phd1 ($P = 0.00026$).

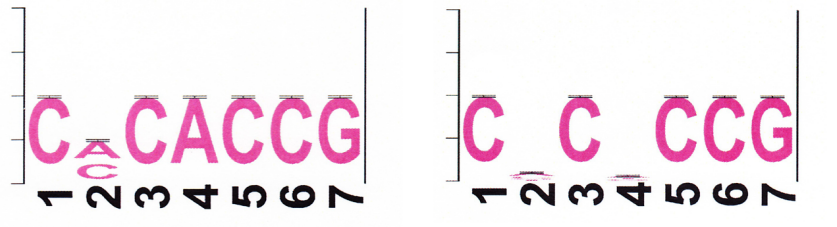
Stp2 interacts with the motif

We hypothesized that Sfp1, Stp2, Phd1, or a combination of these three proteins binds the motif. We inserted a 31-bp sequence from the *AGP2* promoter

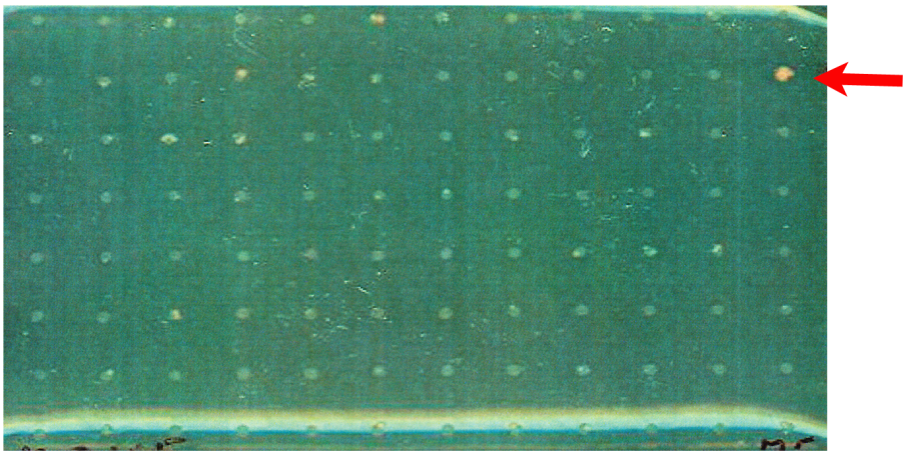
containing two conserved instances of the motif in opposite orientations upstream of a *HIS3* reporter gene. Only the strain carrying Stp2-AD yielded His^+ diploids (Figure 2_3B). Mutations introduced into the first putative binding site in the reporter gene abolish the His^+ phenotype (Figure 2_3C). Mutations in the other binding site significantly diminish the His^+ phenotype. These results suggest that only Stp2 binds to the motif, and not Sfp1 or Phd1.

Genes that were down-regulated in an *stp2*Δ strain and up-regulated in the *STP2* overexpression strain are significantly enriched for the presence of our motif in their promoters ($P = 1.57 \times 10^{-6}$), suggesting that Stp2 is a transcriptional activator that acts through the motif. Stp2 binds specifically to the sequence motif in electrophoretic mobility shift assays (EMSA) using whole-cell extracts from *stp2*Δ, wild-type, and *STP2* overexpressing cells. This DNA–protein complex is super-shifted upon incubation with an antibody specific to overexpressed *STP2*. A twofold excess of the unlabeled sequence motif abolishes the DNA-protein complex while binding is still detected in the presence of a fourfold excess of unlabeled double mutant probe (61).

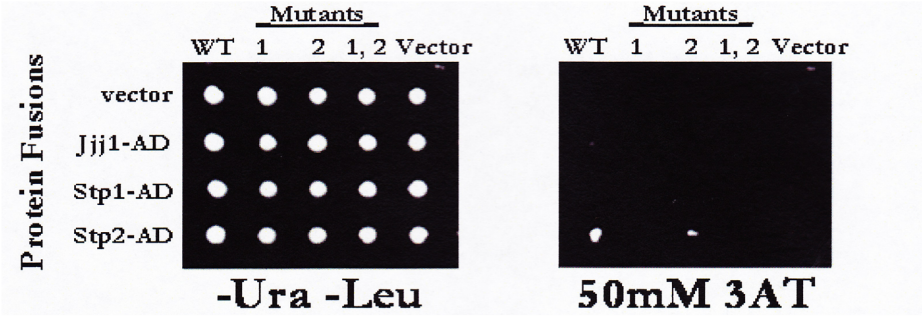
A)



B)



C)



D)

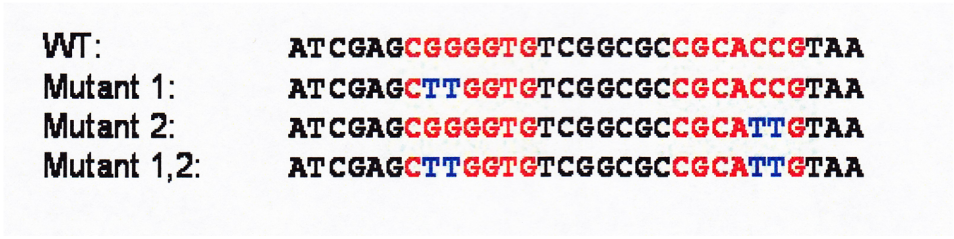


Figure 2_3: Discovery of Stp2 binding site by MY1H

Figure 2_3: Discovery of Stp2 binding site by MY1H

(A) Sequence logos for the putative binding site. **(B)** Ninety-six AD fusions mated to *HIS3* reporter plasmid strains grown on 75 mM 3AT. The arrow points to the Stp2-AD fusion. **(C)** AD fusions mated to mutant versions of a *HIS3* reporter plasmid. **(D)** Sequences used as promoters in one-hybrid assay. Red indicates the motif, and blue indicates mutations.

Discovery of the Sef1 binding site by MY1H

The motifs

75 functionally conserved DNA elements were identified by comparative sequence analysis (31, 32), It was hypothesized that these elements might prove to be binding sites for transcription factors. These 75 motifs were represented in a total of 40 DNA fragments, while 40 additional DNA fragments with mutations in key nucleotides were used as controls (Supplemental Table 3). These DNA fragments are similar to the ones used in the protein microarray experiments (Chapter 3).

We were able to successfully insert 48 of the 80 total DNA fragments into the promoter construct. Of these 48 fragments, 4 showed strong activation and 1 showed strong repression in the absence of any protein-AD fusions.

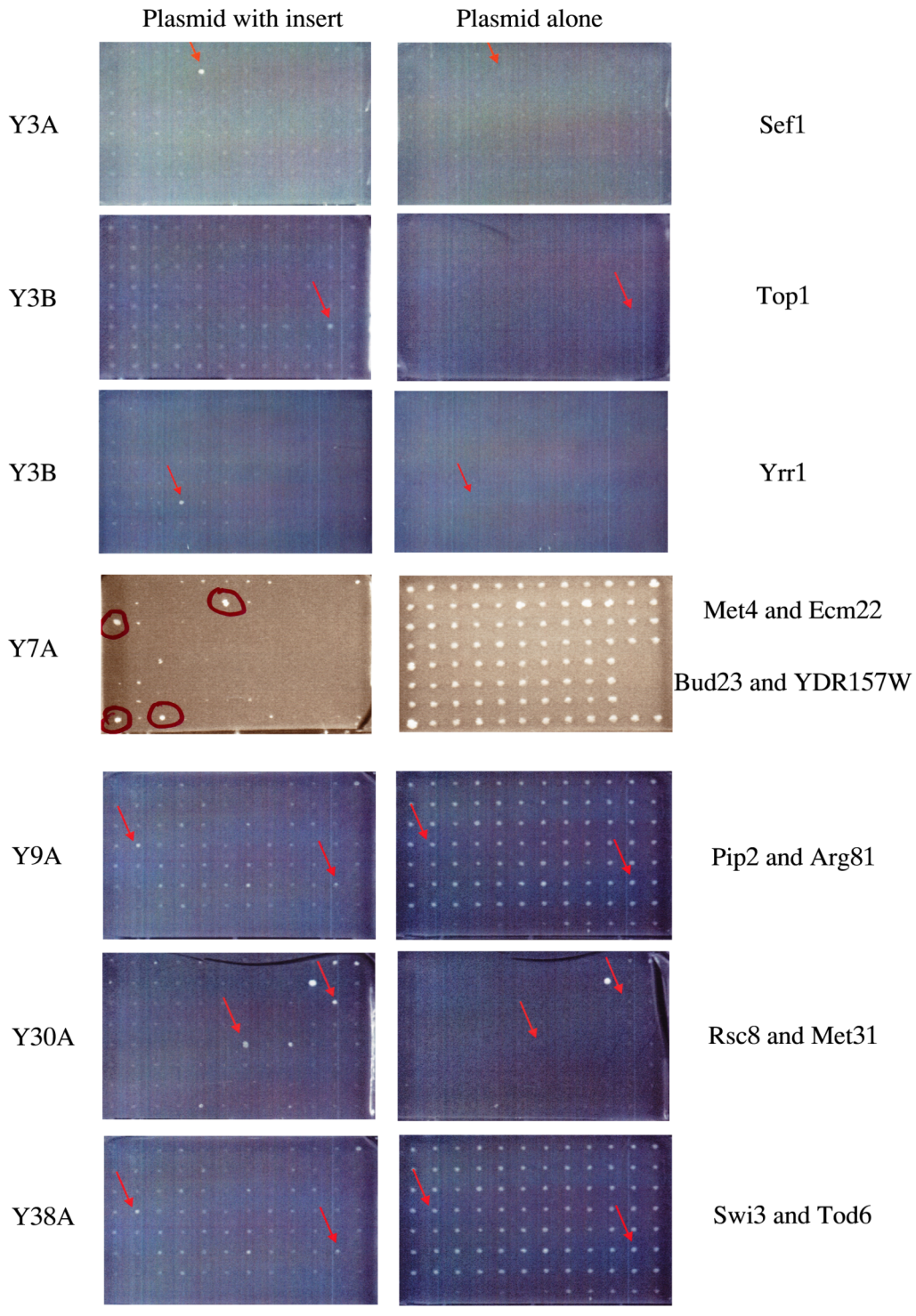


Figure 2_4: MY1H of 13 proteins that bound to DNA elements

Figure 2_4: MY1H of 13 proteins that bound to DNA elements

96 protein-AD fusion strains mated to a strain containing the promoter with the motif, or a promoter alone. Arrows and circles show strains that carry protein-AD fusions that activate the reporter. Each promoter construct required its own conditions: Y3A grown 2 days with 5mM 3AT; Y3B with Top1 grown for 4 days with 20mM 3AT; Y3B with Yrr1 grown for 3 days with 50mM 3AT; Y7A grown for 11days with 20mM 3AT; Y9A grown for 3 days on –his; P30A and Y38A grown for 7 days on 5mM 3AT.

After several rounds of stringent testing, we identified 13 proteins that bound reproducibly to 6 of our DNA fragments (Figure 2_4). 6 of the 13 proteins had binding sites previously assigned to them (8, 78, 133, 265). Many of these assigned sites were computationally derived, and there are discrepancies between sites identified by different methods, for example, four widely different sites have been assigned to Ecm22, a member of the zinc cluster family of transcription factors by four different methods (8, 133, 235, 265). However, only two of these assigned sites contain the canonical CGG triplet that most zinc cluster proteins bind to. Indeed, these two computationally derived sites do not have much in common besides the CGG repeat. In our experiments, the Ecm22-AD fusion bound to a sequence that contained $CGGN_5CGG$, similar to one of the computationally derived sites (265). Yrr1, Met31 and Met4 fusion proteins also bound to inserted motifs that were similar to previously assigned binding sites. It is interesting to note that both Met31 and Met4 have assigned sites based on experimental data, and that the sites in our promoter

constructs were more similar to those sites than to the computationally derived ones.

The sequences that the Arg81 and Tod6 fusion proteins bound to in our MY1H were not similar to their assigned binding sites. However, the binding sites assigned for Arg81 (78) and Tod6 (265) were both computationally derived, and as yet, have not been validated by experimental methods. We are supremely confident that we were able to identify binding sites for at least four of these six proteins with previously assigned binding sites.

Three of the remaining seven proteins function in complexes. Rsc8 is a component of the RSC chromatin remodeling complex; Pip2 forms a heterodimer with Oaf1, and Swi3 is part of the Swi/Snf complex. While this does not preclude them from binding independently to our promoter constructs to an as yet unassigned binding site, our promoter constructs did not contain sequences similar to the binding sites of these complexes. Top1 and Bud23 are both enzymes, and it is possible, though unlikely, that either of these proteins function as transcription factors by binding DNA. YDR157 has since been designated as a dubious ORF. It is likely that these represent false positives in our study.

The last remaining protein is Sef1. No bindings sites have been assigned to this protein. We decided to study this protein-DNA interaction in detail.

Sef1 interacts with Y3A.

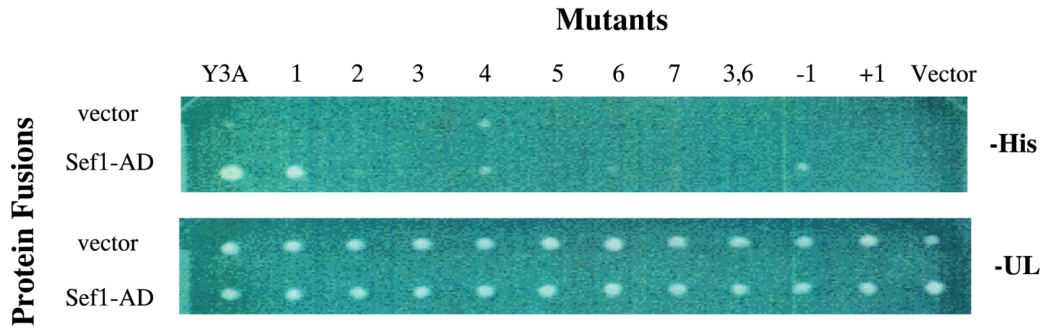
Sef1 is a member of the zinc cluster family of transcription factors. Several members of this family, which includes Gal4, have been well characterized (80, 126). All defined binding sites for this family of proteins consist of CGG triplet, with the

recognition sequences for each protein differing in the orientation (direct, inverted, or convergent) and spacing of the CGG sequences. The Sef1-AD fusion bound to Y3A that contains two copies of the sequence CTACCTCCGAGCTACCTCCGAG. This sequence contains a direct repeat of CGG separated by 8 nucleotides ($CGGN_8CGG$). None of the other 43 DNA fragments tested in this study contained this $CGGN_8CGG$ sequence.

Sef1 binds specifically to $CGGN_8CGG$.

We created several variants of Y3A in order to elucidate the exact sequence that Sef1 bound to. Removal of sequences outside the $CGGN_8CGG$ did not affect binding of Sef1-AD to the promoter. Single point mutations of each of the CGG nucleotides abolished binding, as did a double mutation. A deletion in the sequence leading to $CGGN_7CGG$ reduced binding significantly (Figure 2_5). This indicates that Sef1 binds specifically to $CGGN_8CGG$. Further characterization of Sef1 binding and the role of Sef1 is discussed in Chapter 4.

A)



B)

Y3A:	**GGCTTCTACCTCCGAGCTACCTCCGAGGGATCC**
Mutant 1:	**GGCTTCCGAGCTACCTCCGGGATCC**
Mutant 2:	**GGCTTACGAGCTACCTCCGGGATCC**
Mutant 3:	**GGCTTCAGAGCTACCTCCGGGATCC**
Mutant 4:	**GGCTTCCCAGCTACCTCCGGGATCC**
Mutant 5:	**GGCTTCCGAGCTACCTACGGGATCC**
Mutant 6:	**GGCTTCCGAGCTACCTCAGGGATCC**
Mutant 7:	**GGCTTCCGAGCTACCTCCCGGATCC**
Mutant 3,6:	**GGCTTCAGAGCTACCTCAGGGATCC**
Mutant -1:	**GGCTTCCGAGCTCCTCCGGGATCC**
Mutant +1:	**GGCTTCCGAGCTAGCCTCCGGGATCC**

Figure 2_5: Sef1-AD fusion binds to Y3A but not mutant versions of the sequence.

(A)AD fusions mated to mutant versions of a *HIS3* reporter plasmid. **(B)** Sequences used as promoters in one-hybrid assay. Red indicates the motif, and blue indicates mutations.

Discussion

The yeast one-hybrid assay is one of the many tools in the arsenal to identify the proteins that bind to specific DNA sequences. It has been used in several high-throughput experiments to try to find all binding sites for transcription factors in *S. cerevisiae* (130), and several variations of the system have been created to improve its function (28, 42, 147). Yet despite our best efforts, many binding sites for yeast transcription factors are still uncharacterized. The number of computationally derived binding sites for these transcription factors proliferates (8, 31, 106, 133, 265), yet there are still very few cases in which these sites have been experimentally assigned to specific proteins.

Our modified yeast one-hybrid improves on previous versions using a zero background reporter. In our pilot studies, we did not find many DNA-protein interactions, indicating that our false negative rate is very high. However, this gives us greater confidence of the DNA-protein interactions that we do find. We were able to detect specific DNA-protein interactions of at least four proteins with previously assigned DNA binding sites.

We were also able to identify two novel DNA-protein interactions to two computationally derived DNA sequence motifs. We were able to hypothesize that Stp2 may bind to our motif based on overlap with previous ChIP experiments (122). Sef1 is a putative transcription factor based on homology to a *K. lactis* transcription factor (71), and as yet has not had any binding sites assigned to it by various high throughput computational methods. This emphasizes the utility of our MY1H as no

preconceptions are necessary, and many computationally derived DNA sequence motifs can be quickly tested to find positive interactions.

CHAPTER 3

Linking DNA-binding proteins to their recognition sequences by using protein microarrays

(published in Proceedings of the National Academy of Science,
2006 Jun 27; 103(26):9940-5)

Abstract

Analyses of whole genome sequences and experimental datasets have revealed a large number of DNA sequence motifs that are conserved in many species and may be functional. However, methods of sufficient scale to explore the roles of these elements are lacking. We describe the use of protein arrays to identify proteins that bind to DNA sequences of interest. A microarray of 282 known and potential yeast transcription factors was produced and probed with oligonucleotides of evolutionarily conserved sequences that are potentially functional. Transcription factors that bound to specific DNA sequences were identified. One previously uncharacterized DNA-binding protein, Yjl103, was characterized in detail. We defined the binding site for this protein and identified a number of its target genes, many of which are involved in stress response and oxidative phosphorylation. Protein microarrays offer a novel high-throughput method for determining DNA-protein interactions.

Introduction

A fundamental problem in biology is to identify *cis*-regulatory DNA sequence elements and the proteins that bind to them. Such information is necessary for uncovering gene regulatory networks that control cellular and developmental processes. Genome-wide approaches have revealed many DNA sequence elements that may regulate gene expression: comparison of genome sequences of related organisms has identified thousands of evolutionarily conserved DNA sequence motifs (31, 106, 195); comparison of the sequences adjacent to co-regulated sets of genes of an organism often reveals shared sequence-motifs (31, 87, 212, 232). Verifying functionality of these sequences and identifying the proteins that bind to them remains a significant challenge.

Several methods have recently been developed to map globally the DNA-binding sites of transcription factors. The SELEX method enables *in vitro* selection of the optimal binding site of a transcription factor (194), though applying it genome-wide may be difficult. In the "ChIP-chip" method, chromatin bound by a transcription factor of interest is immunoprecipitated and the associated DNA is identified by using it to probe a genomic DNA microarray, thereby identifying the targets of the transcription factor (85, 189). Two related methods are direct probing of a DNA microarray with a DNA-binding protein, or capture of genomic DNA *in vitro* with a DNA-binding protein, followed by its identification by probing a DNA microarray ("DIP chip") (127, 156). While these methods have achieved considerable success, their resolution is comparatively low because they identify relatively large segments of DNA bound by a protein. Pinpointing the binding site within these

segments requires inference (usually by computational analysis). Indeed, the DNA sequences recognized by over half of the predicted DNA-binding proteins in yeast remain to be identified.

Although these methods promise comprehensive identification of the targets of a known transcription factor, they are not able to do the converse: identify the binding protein that recognizes a sequence motif of interest. Thus, they are unable to take advantage of the thousands of conserved functional DNA sequence elements that have been predicted from a variety of studies and whose DNA-binding proteins are unknown (32, 61, 87, 106, 232, 245). One method that potentially offers this capability is the one hybrid method for identifying proteins that bind to a particular sequence *in vivo* (42), but its application on the whole genome scale may be difficult.

To fill this void, we have developed a novel high-throughput method for identifying sequences recognized by DNA-binding proteins that employs an array of transcription factors. Oligonucleotides containing evolutionarily conserved DNA sequence motifs were used to probe an array of approximately 300 known or potential transcription factors from *S. cerevisiae*. We identified numerous protein-DNA interactions, and characterized the DNA sequence recognized by a previously uncharacterized DNA binding protein. This method should be applicable to any organism.

Results

Development of protein arrays for assaying DNA-binding activity.

We first tested if proteins arrayed on a surface could be used to detect specific protein-DNA interactions by arraying a few transcription factors (Rap1, Abf1, Swi6) whose binding sites are well-defined, along with two proteins that do not bind to DNA (Cmd1p, and Cmk1p). This mini array was probed with a Cy3-labeled oligonucleotide containing three copies of the canonical binding site of Rap1, prepared as described in Fig. 3_1A. Multiple copies of the Rap1 recognition sequence were incorporated into the probe to increase the local concentration of binding sites. A Cy5 labeled probe with two base-pair changes in the central invariant nucleotides of the binding sites was used in parallel to test the specificity of binding (Fig. 3_1B; see Materials and Methods).

The proteins were arrayed on a variety of different surfaces and probed under different conditions (see Materials and Methods). Conditions were identified in which Rap1 bound to the "wild-type" probe but not to the "mutant" probe, regardless of the fluorophore used to label the probes (Fig. 3_1B). These probes did not bind to any other DNA-binding proteins on the array, or to the non-DNA-binding proteins, indicating that binding is specific. In all of our preliminary experiments we tested a total of 7 proteins with binding sites of known sequence: Rap1, Zap1, Ume6, Yap1, Abf1, Swi6 and Mbp1. The first 5 of these proteins bound to probes containing their known binding sites (Fig. 3_1, and data not shown.

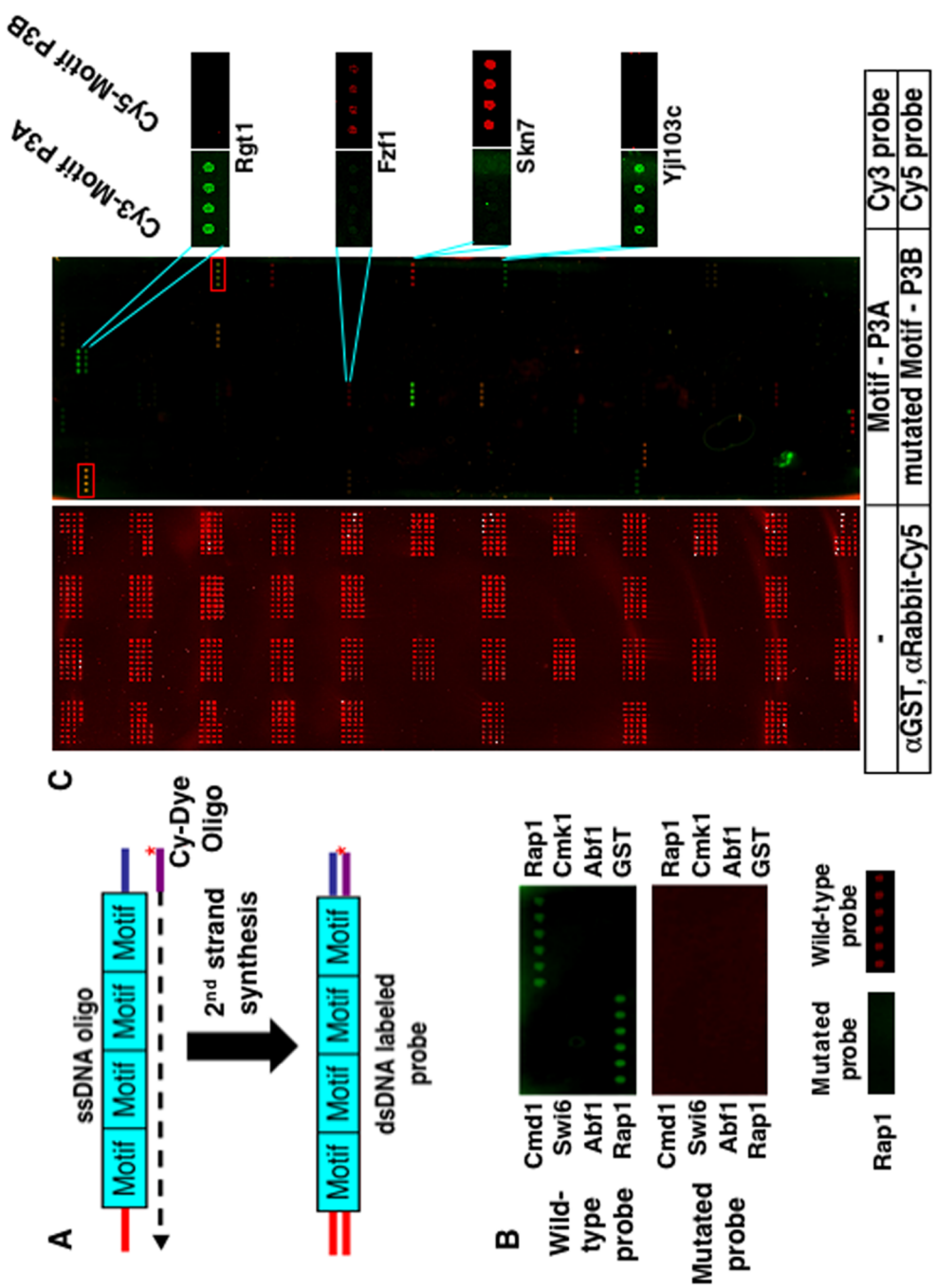


Fig. 3_1: Probing the transcription factor microarray.

Fig. 3_1: Probing the transcription factor microarray. (**A**) Probes were made by extending a universal primer labeled at its 5' end with a fluorophore on an oligonucleotide template containing conserved sequence motifs. Because the length of the sequence motifs vary and we kept the length of the oligonucleotide probes constant, 3 or 4 copies of a motif are present in each probe. **(B)** Rap1 protein binds to a probe containing Rap1 binding sites. Each protein depicted on the right and on the left of the panels was spotted 6 times on the nitrocellulose surface and probed with an oligonucleotide containing three Rap1 binding sites (ACACCCAT/GCA) (labeled with Cy3, shown in green) and a probe containing three Rap1 binding sites with 2 nucleotide changes (ACACttAT/GCA) (labeled with Cy5, shown in red). Probing with reciprocally labeled probes is depicted in the bottom of the panel (**C**) Yeast transcription factor microarrays probed with fluorescent DNA probes. The GST-fused transcription factors purified from yeast (see Materials and Methods) were spotted (in quadruplicate) on each slide and probed with Cy5-labeled anti-GST (left panel), or a pair of probes (right panels). Examples of specific DNA binding are enlarged at the right. Yjl103c binds specifically to P3A but not P3B.

Surveying proteins that bind to conserved sequences using a transcription factor array.

To identify proteins that bind to specific DNA sequence motifs, we produced a microarray of two hundred eighty-two known or potential DNA-binding proteins chosen based on their GO designation as transcription factors, their homology to known DNA-binding domains, or their association with an *in vitro* DNA binding

activity (77). Most of the proteins known to bind DNA non-specifically, such as chromatin binding proteins and subunits of the general transcription machinery, were excluded from the array. The proteins were expressed in yeast cells as fusions to Glutathione-S-transferase (GST), purified by glutathione affinity chromatography (265), and spotted on microscope slides (Fig 3_1C). The concentration of protein in each spot varied from approximately 0.2 to 4 ng/μl.

The transcription factor array was probed with 40 Cy3-labeled double stranded DNA oligonucleotides containing, in total, 75 novel DNA sequence motifs previously identified by their evolutionary conservation (31) (Supplemental Table 4). Each oligonucleotide probe contained 3 or 4 copies of the sequence motif to be tested (Fig. 3_1A). We were able to represent the 75 sequence motifs in 40 oligonucleotides by careful design of the junctions between the repeated sequence motifs. To distinguish between specific DNA-protein interactions and non-specific interactions, the array was probed with a second set of "mutant" Cy5-labeled probes that contained two base-pair changes in the conserved sequence motifs (Fig. 3_1C) (Supplemental Table 4). Since sequence motifs are relatively short, judicious design of the "mutant" probes meant that two base-pair changes in each copy of the sequence motif changed the sequence of most motifs represented in the oligonucleotide. This also creates novel sequence motifs absent in the "wild-type" probes. For example, Fzf1, which recognizes TATCGTAT (6), binds to the two "mutant" probes (P3B (Fig. 3_1C) and P30B (Supplemental Table 4)), because they contain the sequences TATCG and

TATGGTGT. These sequences are not represented in the corresponding probes (P3A and P30A) that serve as the variants of the P3B and P30B probes.

Twenty-three proteins on the array appear to bind DNA non-specifically because they bound to most probes (Fig. 3_1C) with approximately equal affinity (shown in yellow), and to a double-stranded oligonucleotide consisting of the universal sequences that flank the motifs in each probe. These proteins, which included several known non-specific DNA-binding proteins, such as Nph6A/B and Htz1, were excluded from further analysis. Some proteins bind more strongly to the "wild-type" DNA probe (shown in green); others bind more strongly to the "mutant probe" (shown in red). Sixty-two proteins on the array bound to at least one probe. (Supplemental Table 5)

Many Specific DNA-Protein Interactions Can Be Detected

We identified a total of 211 specific DNA-protein interactions with the 80 probes (40 pairs of probes). Thirty-five probes did not interact specifically with any proteins on the array; 9 probes had only one specific DNA-protein interaction; 15 probes bound to between 5 and 22 different proteins. This latter result is not surprising since concatenation of motifs creates multiple binding sites that can be recognized by different proteins.

Among the 211 specific DNA-protein interactions detected, 80 involved proteins with previously characterized binding sites (30 total proteins), including

Met31 and Met32, which have been shown to bind as a heterodimer (15). For 17 of these proteins their characterized/canonical binding sites are present in at least one bound probe. This is a minimal positive rate (17/30 or 57%) because we avoided including known binding sites in the probes as much as possible. Surprisingly, the putative recognition sequence was not apparent in the probes that bound to 13 previously characterized proteins. Perhaps the sequences recognized by these proteins are not well defined, or perhaps they recognize more than one sequence.

We further analyzed 8 of these proteins whose DNA-binding sites are not known: Yjl103, Rgm1, Ypr196, and Rds2, each of which bound a single probe; Stp4 which bound 2 probes; Stp3 and Hms1, which bound 4 probes; Yml081, which bound 5 probes. Stp3 and Yml081 bound to probes containing sequences similar to their respective binding sites predicted using the model described by Benos et al.(14) (G. Stormo, personal communication).

To verify the specific DNA interactions of the 8 proteins, we purified the proteins, incubated them with their corresponding probes and subjected them to electrophoretic mobility shift assays (EMSA). We were able to detect specific binding of the appropriate probes to 7 of the 8 proteins (Fig. 3_2), confirming that their binding sites are contained within the probe sequence.

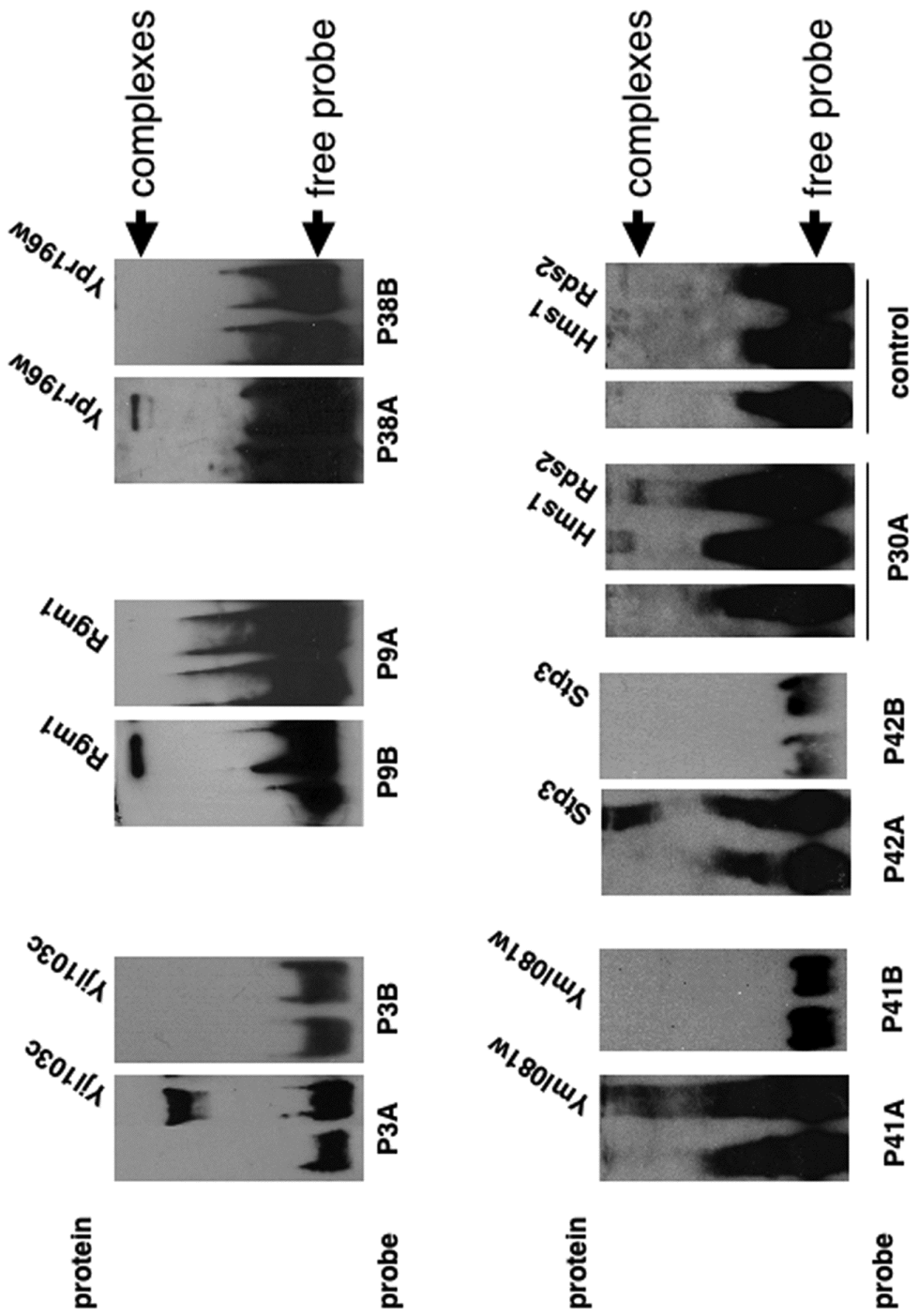


Fig 3_2: Electrophoretic Mobility Shift Assays (EMSA) of 7 proteins that showed specific DNA-binding on protein microarrays.

Fig 3_2: Electrophoretic Mobility Shift Assays (EMSA) of 7 proteins that showed specific DNA-binding on protein microarrays. Only one probe of each probe pair (left panel) binds specifically to the protein. There are two or three base pair differences in each motif in each pair of probes. P30A binds to both Hms1 and Rds2. P38A is used as a control to show binding is specific to P30A. See Materials and Methods for details.

Yjl103 binds to $CGGN_8CGG$.

One protein-DNA interaction was studied in detail. Yjl103 is a member of the zinc cluster family of transcription factors. Several members of this family, which includes Gal4, have been well characterized (80, 126). All defined binding sites for this family of proteins consist of CGG repeats, with the recognition sequences for each protein differing in the orientation (direct, inverted, or convergent) and spacing of the CGG sequences. Yjl103 binds to a probe containing two overlapping copies of a direct repeat of CGG separated by 8 nucleotides ($CGGN_8CGG$). None of the other 39 probe pairs contain this sequence.

Yjl103 binds to its specific probe and not to the mutant probe in gel shift assays (Fig. 3_3). Binding was competed with a sequence containing a single copy of $CGGN_8CGG$; an oligonucleotide containing the sequence $CTGN_8CTG$ did not compete for binding.

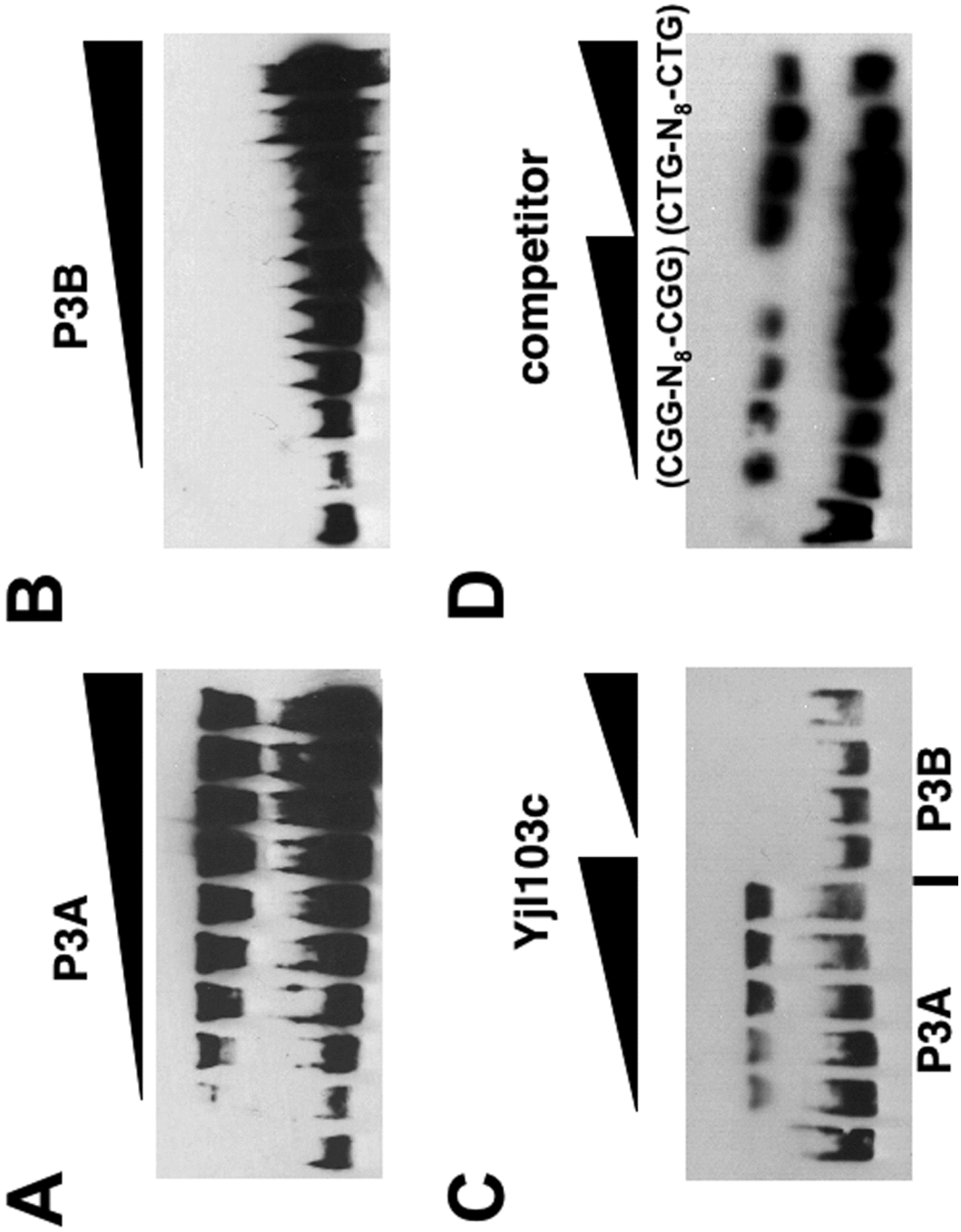


Fig. 3_3: EMSA of Yjl103c.

Fig. 3_3: EMSA of Yjl103c. (**A** and **B**) A constant amount of Yjl103c (5μM) incubated with increasing amounts of labeled probes P3A ("wild-type" binding site) and P3B ("mutant" binding site) respectively. Probe concentrations increase from 60 ρM to 600 ρM. (**C**) Constant concentration of probes P3A and P3B (250 ρM) with increasing amounts of Yjl103c. Protein increases from 0.7 μM to 8.5 μM. (**D**) Competition with unlabeled DNA: An increasing amount of cold competitor DNA is added to the reaction with constant concentration of Yjl103c (1.6μM) and labeled probe P3A (250 ρM). Cold competitor is added at effective excess of labeled probe of 10 fold, 50 fold, 100 fold and 800 fold.

The protein chip assay was used to further elucidate the binding site for Yjl103. The inclusion of metal chelators (EGTA and especially EDTA) during the probing severely impaired the binding of Yjl103 to DNA, suggesting that zinc is important for its DNA binding activity (data not shown). A Yjl103-GST fusion protein was purified from yeast, immobilized on a surface and incubated with a panel of probes containing variants of the $CGGN_8CGG$ sequence (Fig. 3_4). The first CGG appears to be required for binding, but the latter two residues in the second CGG appear to be less important because substitutions in either of these positions reduce, but do not abolish binding. Yjl103 binds *in vitro* to both $CGGN_8CGG$ and $CGGN_9CGG$. This is somewhat surprising because other members of this family of DNA-binding proteins appear to have a strict requirement for a specific spacing of the CGG repeats.

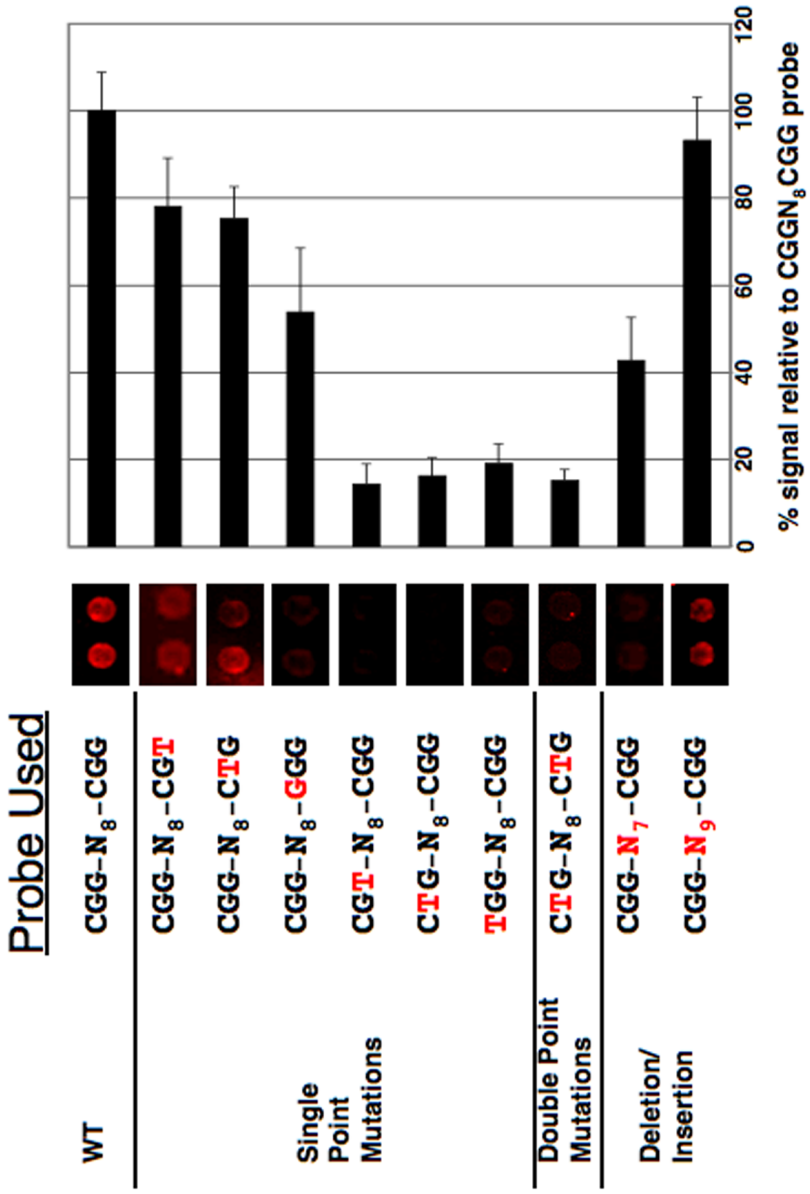


Fig. 3_4: Yjl103c binds to $CGGN_8CGG$ and $CGGN_9CGG$.

Fig. 3_4: Yjl103c binds to $CGGN_8CGG$ and $CGGN_9CGG$. Oligonucleotides containing variations of the putative binding site of Yjl103c were used to probe the transcription factor microarrays. Binding intensity, relative to the "wild-type" probe, is plotted on the right (average of 3-5 independent probings with each sequence)

Yjl103 binds upstream of genes with $CGGN_{8/9}CGG$ that are involved in energy utilization.

To identify targets of Yjl103 and thereby gain clues to its function, we compared the gene expression profile of a wild-type strain to those of strains that over express or are deleted for *YJL103C*. Over 500 genes were differentially expressed between the wild-type and the *YJL103C*-overexpressing strains (about half of these were up-regulated by Yjl103 overexpression). These genes are enriched for proteins involved in carbon compound and carbohydrate metabolism ($P = 3.73x10^{-5}$) and also for proteins involved in stress response ($P = 4.79x10^{-5}$), two roles previously suggested for Yjl103 (41). We found 131 genes that were expressed differently in the *yjl103cΔ* mutant compared to the wild-type strain (23 of them were among the 551 genes affected by *YJL103C* overexpression), about two-thirds of which are up-regulated in the deletion mutant. Thirty-five of the more than 500 genes whose expression was altered by *YJL103C* overexpression, and 7 of the 131 genes whose expression was altered by deletion of *YJL103C*, contain $CGGN_8CGG$ or $CGGN_9CGG$ in their promoters. These are not significantly more than expected by chance, but this

may be because we have not yet found the optimal conditions for inducing Yjl103 function.

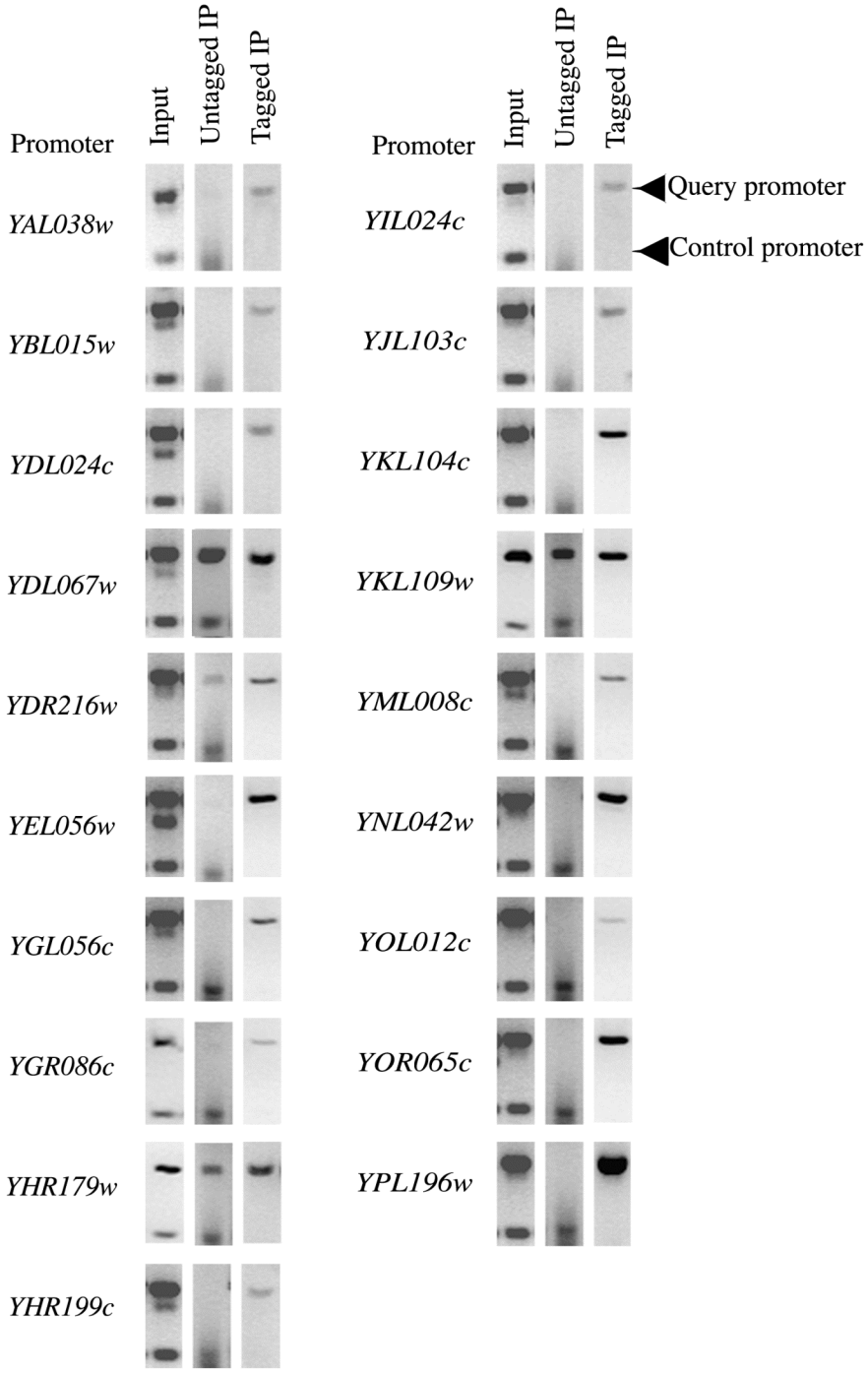


Fig. 3_5: Chromatin immunoprecipitation assay for Yjl103 binding.

Fig. 3_5: Chromatin immunoprecipitation assay for Yjl103 binding. Chromatin was crosslinked to proteins, Yjl103 tagged with a 13-myc epitope was precipitated with anti-myc antibody, and the precipitated DNA was released from protein and detected by a PCR (as described in Materials and Methods) using primers specific for sequences upstream of the indicated 19 genes (query promoter), and primers specific for the *GAL4* promoter (control promoter) that amplify a 150 base-pair fragment.

To determine whether Yjl103 binds *in vivo* to any of these genes whose expression is altered in strains lacking or overexpressing Ygl103 and that contain a $CGGN_8CGG$ sequence motif upstream, several of them were tested for Yjl103 binding using chromatin immunoprecipitation. As shown in Fig 3_5, 19 of 22 genes were enriched in immunoprecipitates from a Yjl103::c-myc tagged strain relative to controls. Thus, Yjl103 associates with $CGGN_8CGG$ targets *in vivo* as well as *in vitro*.

Discussion

Although a large number of potentially functional DNA sequence motifs have been identified from gene expression and sequence conservation studies, no facile method for identifying the proteins that bind to them has been available. Here we describe implementation of protein arrays for this purpose. Using a microarray of yeast transcription factors, we were able to detect many known and new DNA-protein interactions. Nucleotide substitutions in the known binding sites completely abolished binding of a protein, providing validation for the assay. In this way, we were able to

define the proteins that bind to several sequence motifs and discover a previously unknown DNA-binding specificity.

Many of the proteins that bound to our probes bound to many probes, suggesting that they bind to DNA nonspecifically. For example, Phd1 bound 11 probes with no common sequence among them by eye. Although many of these proteins probably bind DNA nonspecifically *in vitro*, such as Nhp6A and Nhp6B, others like Phd1 seem to bind specific sequences *in vivo* (16). It is therefore likely that these proteins use additional cofactors to achieve sequence-specific binding.

The transcription factor arrays were probed with oligonucleotides containing multiple copies of the sequence motifs. Thus, each probe may contain several overlapping binding sites, capable of being recognized by several proteins. The pattern of binding by each transcription factor can often be deconvoluted by examining the different probes each one binds. For example, Hms1 bound three probes, each of which contained the sequence ACCACA. Probes that bound to Yml081 also contained sequences similar to its predicted binding motif (14). In other cases, it is necessary to determine the exact sequence bound by the particular factor. One possible solution to this problem is to separate defined sequence motifs with random nucleotides, which would keep other binding sites at single copy while retaining in the probe multiple copies of the intended binding site (Supplemental Table 4).

Many transcription factors bind to DNA as heterodimers. It is noteworthy that we detected binding of both Met31 and Met32 to a probe containing the binding site of this heterodimer. We expect that other heterodimeric DNA-binding proteins purified from yeast extracts will similarly be associated with their partner protein(s). It should also be possible to carry out the binding reactions in the presence of another protein, or perhaps in the presence of a nuclear extract, to recreate heterodimers not present on the array. Combinations of proteins could also be spotted on the array, providing a matrix of all possible heterodimers.

We characterized in detail the binding site of a previously uncharacterized protein: Yjl103, a member of the Zn cluster family of transcription factors whose bindings sites are variations of CGG repeats. We defined the binding site of Yjl103 to be 2 direct repeats of CGG separated by 8 or 9 nucleotides ($CGGN_8CGG$ or $CGGN_9CGG$). It is somewhat surprising that the spacing of the CGG repeats is variable, since the binding sites of nearly all members of this family of DNA-binding proteins have rigid spacing requirements. In fact, it is the spacing of the CGG repeats (and their orientation) that determines the specificity of DNA-binding of each protein. Perhaps Yjl103 forms a complex with other proteins that modify its sequence spacing requirement. Gene expression profiling identified several genes differentially regulated when Yjl103 is overexpressed or deleted. Yjl103 binds *in vivo* upstream of 19 of 22 of the genes we tested, and all of them contain the $CGGN_8CGG$ sequence motif. The known or predicted functions of the proteins encoded by these genes are

enriched in carbon compound and carbohydrate metabolism, consistent with the proposed role of Yjl103 in energy utilization (41).

In yeast, a very well characterized organism, the sequences recognized by only about half of its 200 or more transcription factors are known. Protein array technology offers the possibility for high-throughput analysis of all transcription factors with many probes under a variety of conditions, and should bring the catalogue of transcription factor binding sites within our reach. Application of this technology to mammals, with approximately 1000-1500 transcription factors, would require only a modest increase in the scale of the analysis. Thus, it should be possible to determine *cis*-regulatory sequences and the proteins that bind to them across the genome, which is the first step in decoding the regulatory networks of an organism.

CHAPTER 4

Characterization of SEF1 and YJL103C

Abstract

I identified the binding sites of Sef1 and Yjl103c using two different high-throughput assays. Both proteins appear to bind to the sequence $CGGN_8CGG$. I further characterized the binding site of Sef1.

Although I was not able to show physical interaction between Sef1 and Yjl103c, I found that binding to $CGGN_8CGG$ is cooperative, and that both proteins regulate some of the same genes. Both Sef1 and Yjl103c associate with proteins involved in glycolysis and gluconeogenesis. Sef1 and Yjl103c function as repressors that appear to have similar but distinct roles, as both regulate *ADR1* expression but only *YJL103C* regulates *SDS23* expression. They also participate in a feedback loop with Hap4, a global regulator of respiratory gene expression. *PHO84* and *SPL2* are highly expressed in *sef1Δ* and *yjl103cΔ* strains, suggesting communication between the internal levels of inorganic phosphate and various metabolitic pathways. I propose that Sef1 and Yjl103c coordinate multiple metabolitic pathways.

Introduction

In the era of high-throughput assays, it is tempting to assume that one will in the near future be able to identify and characterize the entire inner workings of an organism (88). Anecdotal and systematic examination of our knowledge of yeast genes reveals that accumulation of facts does not lead instinctively to universal understanding of gene function. Indeed, two recent high-throughput studies of all yeast transcription factors using Protein Binding Microarrays (PBMs) claim to have identified the sequences recognized by up to 80% of the transcription factors in *S. cerevisiae* (8, 265), but upon closer inspection, some of these sequences bear little resemblance to ones previously identified for some of those proteins. Nevertheless, it appears that individual efforts aimed at understanding the functions of single genes benefit from large-scale research, and the benefit is reciprocal, because gold-standard annotations assist in the interpretation of large-scale data sets.

I developed two high-throughput assays designed to identify proteins that bind to specific DNA sequence motifs. Using these assays, I identified two proteins that bind to the sequence $CGGN_8CGG$. Sef1 was identified using a modified yeast one-hybrid assay; Yjl103c was identified using protein microarrays.

Little is known about these proteins. Sef1 (Suppressor of Essential Function 1) was first identified as a protein that complements the essential function of Rpm2 in *K. lactis* (71), a mitochondrial Rnase P. *RPM2* is essential for growth on glycerol in *K. lactis*, and acts as a transcriptional activator in the nucleus to maintain the steady state mRNA levels of some nuclear-encoded mitochondrial components (216). Although Sef1 is a member of the extensively studied zinc cluster family of transcription

factors of *S. cerevisiae*, it has been strikingly absent from research aimed at elucidating this family of transcription factors, and of transcription factors in *S. cerevisiae* in general. There has been no predicted binding sites or functional characterization of Sef1.

The same cannot be said for Yjl103c. In the past three and a half years since I identified the binding site of Yj1103c using protein microarrays, it has been used in two studies to illustrate the efficacy of annotating unknown yeast ORFs by computational methods using published information such as microarray expression data (41, 178). Both groups hypothesized that Yjl103c encodes a protein involved in regulation of the respiratory pathway in yeast. They gave it the name Gsm1 (Glucose Starvation Modulator 1) (*Saccharomyces* Genome Database; http://www.yeastgenome.org) based on one of these studies (41)- but for the sake of continuity, I will persist in referring to the protein as Yjl103c. Yjl103c has also been used in development of an improvement of the ChIP-chip technique (228), although no biological analysis of the data was presented. Two studies have postulated a binding site similar to the one I identified, but the proposed binding site contains a single CGG (8, 265). However, it appears that data largely captures monomeric specificities rather than the dimeric motifs typically associated with zinc cluster proteins (8).

The information encoded in a gene promoter is decoded primarily by the sequence specific binding of transcription factors (62). Thermodynamic modeling of synthetic promoters shows that some binding sites switch modes of action due to the competition between multiple factors for the same site and it is likely that this mode

of regulation allows for drastic changes in expression in response to changing transcription factor concentrations (60). For example, it has been proposed that Nrg2, a protein similar to Nrg1, binds to the same site as Nrg1, and both may be phosphorylated by the Snf1 protein kinase. (239) .

I showed that Sef1 and Yjl103c bind to exactly the same sequence *in vivo* and *in vitro*. This led me to hypothesize that Sef1 and Yjl103c have similar roles in the process of energy utilization. Both proteins regulate expression of the transcription factor genes *ADR1* and *HAP4*, which are involved in energy utilization. The roles of Sef1 and Yjl103c appear similar, yet distinct, as they regulate a subset of the same genes, but have some dissimilar targets. They are also regulated differently. I propose that Sef1 and Yjl103c play a role in coordinating multiple nonfermentable metabolic pathways in *S. cerevisiae.*

Results

Sef1 and Yjl103c bind to the same sequence *in vitro* and *in vivo.*

SEF1 encodes a 1,148 amino acid protein; *YJL103C* encodes a protein about half that size (618 amino acids). I was unable to find any similarity between the two proteins besides the canonical consensus sequence $CysX_2CysX_6CysX_{5\text{-}12}CysX_2CysX_{6\text{-}8}Cys$ (Figure 4_1A) of the zinc cluster transcription factors in the N-terminal regions. No other recognizable domains were found in either protein, although there are claims that Yjl103c contains a characteristic cytochrome C signature, which is a cytochrome C family heme-binding site (41). However, this sequence is within the zinc finger of Yjl103c, which would seem to preclude Yjl103c being a heme-binding protein.

I identified the binding site of Sef1 from the MY1H assay (Chapter 2). Yjl103c was not identified in the screen, even though the protein was present in the library of transcription factor-AD fusions I used, but I was able to demonstrate binding of the Yjl103c-AD fusion to the same DNA sequence (Y3A) used in Chapter 2 to identify Sef1 (Figure 4_1B). This binding is specific, as the Yjl103c-AD fusion did not bind to a DNA sequence (Y3B) that does not include $CGGN_8CGG$. In a dilution series using either the Sef1-AD fusion or the Yjl103c-AD fusion, transcriptional activation by the Sef1-AD fusion was at least twenty times stronger than by Yjl103c-AD (Figure 4_1C). This explains why Yjl103c was not identified in the MY1H screen.

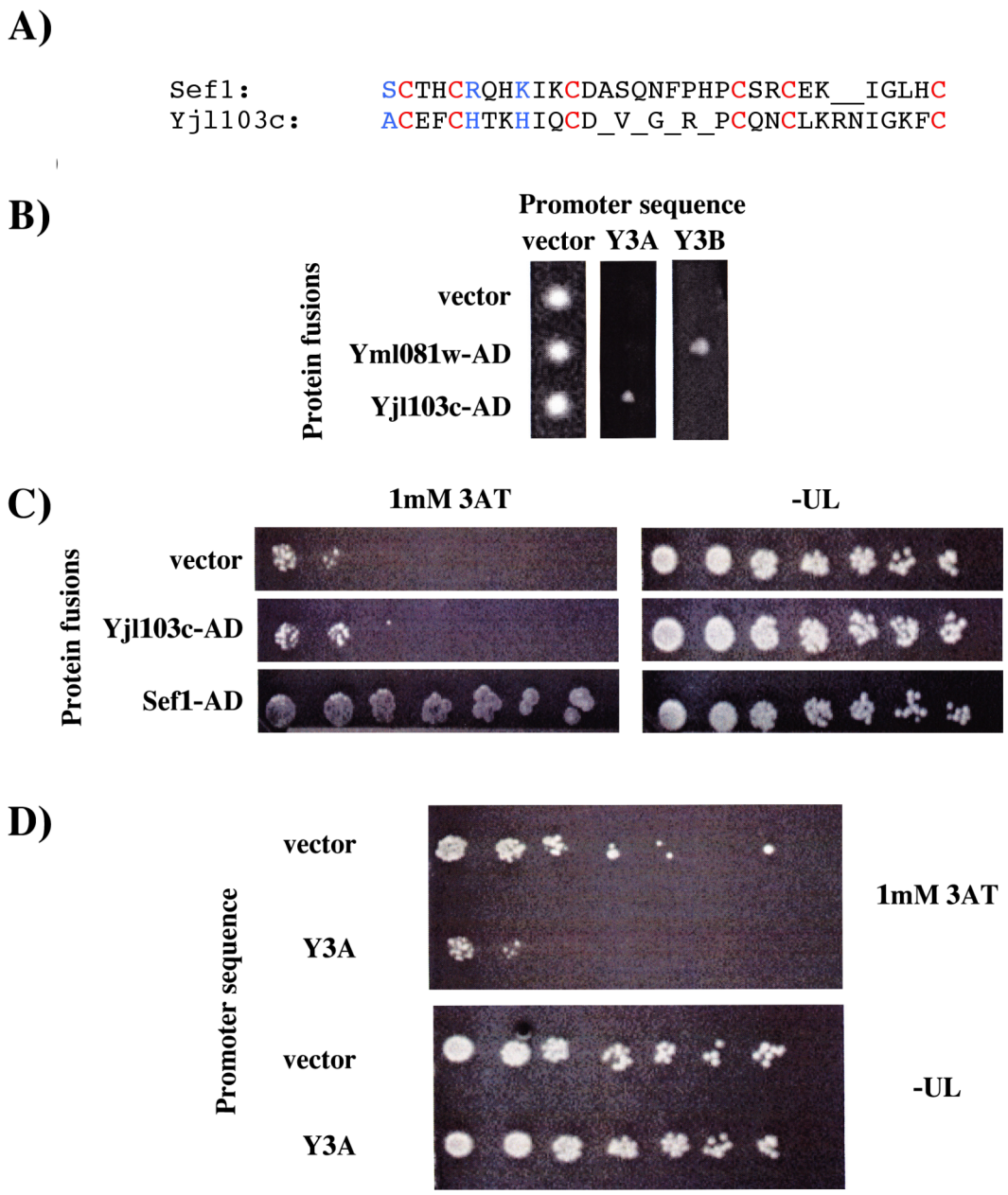
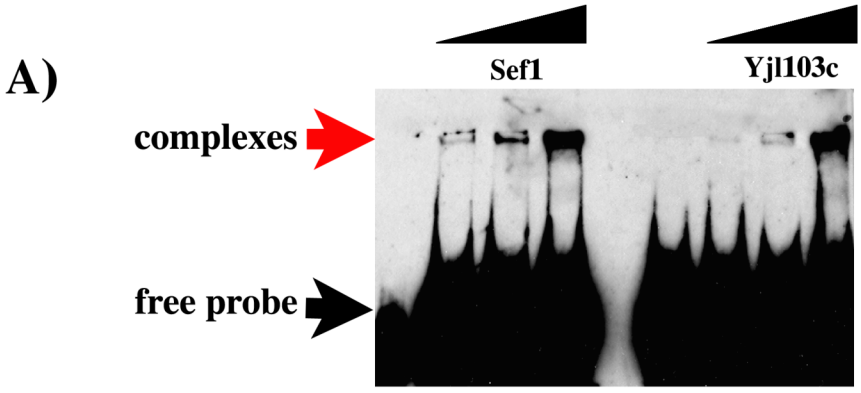


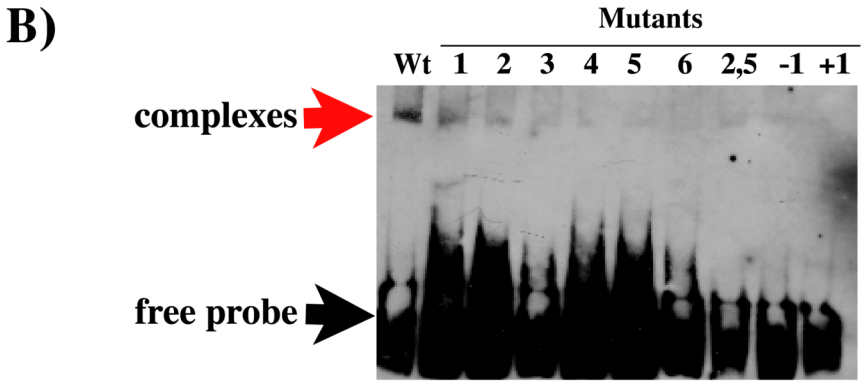
Figure 4_1: MY1H with Sef1 and Yjl103c

(A) Alignment of consensus Zn fingers of Sef1 and Yjl103c. In blue are characteristic amino acids, and in red are the cysteines. **(B)** MY1H of Yjl103c-AD with DNA sequences Y3A and Y3B. Strains containing the promoter with Y3A were grown for

14 days on 20mM 3AT and strains containing the promoter with Y3B were grown for 14 days on 50mM 3AT. The vector alone control strains were grown on –ura –leu media. **(C)** and **(D)** Titration of cells. Cells were diluted sequentially 1:5 and grown for 4 days on 1mM 3T. Number of cells in each strain are comparable as seen in by growth of controls on –ura –leu media **(C)** Cells expressing Yjl103c-AD or Sef1-AD fusions in the presence of the promoter containing Y3A. **(D)** Cells containing the vector alone or the promoter with Y3A in the absence of any AD fusion proteins.

Sef1 was present on our protein microarrays (Chapter 3) but I discovered that the version used was truncated (data not shown). I used a full-length Sef1-GST fusion protein in an EMSA, and found that Sef1 binds to the same sequence that Yjl103c bound to on the protein microarrays (oligonucleotide P3A, Figure 4_2A). To further characterize the binding site of Sef1, I assembled a panel of probes containing variants of the $CGGN_8CGG$ sequence (Figure 3_4) (Figure 4_2B). The binding of Sef1 to the mutant probes recapitulated binding to mutant sequences in the MY1H (Figure 2_5A). As with the MY1H, the first CGG appears to be required for binding but the last residue in the CGG seems to be less important, as substitution of this residue does not abolish binding. This is also the residue of least importance for Yjl103c binding (Figure 3_4). In contrast to Yjl103c, which appears to bind both $CGGN_8CGG$ and $CGGN_9CGG$ *in vitro*, Sef1 appears to bind $CGGN_8CGG$ and to a lesser extent, $CGGN_7CGG$ both *in vivo* and *in vitro*.





C)

Wt:	TCGACTCTGGATCCCCGAGCTACCTCCGCATAGTCTACTA
Mutant 1:	TCGACTCTGGATCCACGAGCTACCTCCGCATAGTCTACTA
Mutant 2:	TCGACTCTGGATCCCAGAGCTACCTCCGCATAGTCTACTA
Mutant 3:	TCGACTCTGGATCCCCCAGCTACCTCCGCATAGTCTACTA
Mutant 4:	TCGACTCTGGATCCCCGAGCTACCTACGCATAGTCTACTA
Mutant 5:	TCGACTCTGGATCCCCGAGCTACCTCAGCATAGTCTACTA
Mutant 6:	TCGACTCTGGATCCCCGAGCTACCTCCCCATAGTCTACTA
Mutant 2,5:	TCGACTCTGGATCCCAGAGCTACCTCAGCATAGTCTACTA
Mutant -1:	TCGACTCTGGATCCCCGAGCTCCTCCGCATAGTCTACTA
Mutant +1:	TCGACTCTGGATCCCCGAGCTAACCTCCGCATAGTCTACTA

Figure 4_2:EMSA of Sef1 and Yjl103c with labeled probes.

(A) Constant concentration of probe P3A (250ρM) with increasing amounts of Sef1 or Yjl103c. Protein increases from 0.7μM to 8.5μM. As the amount of protein in the reaction increases, the band corresponding to the shifted protein-DNA complex increases in intensity (red arrow). **(B)** Sef1 is incubated with variants of

$CGGN_8CGG$. Protein concentration is kept constant at 4μM. The band due to the DNA-protein complex (red arrow) is strongest in the wild-type probe lane. Some DNA-protein complex is found in mutant 1. **(C)** Sequences used as probes in (B). Red indicates the motif, and blue indicates mutations.

Sef1 and Yjl103c function as repressors of transcription.

Figure 4_1D shows that insertion of the DNA sequence GGCTTCTACCTCCGAGTACCTCCGAGGGATCC (Y3A) into the promoter construct used in the MY1H actually reduces base-line expression of *HIS3* in the absence of any protein-AD fusion. This led me to hypothesize that endogenous proteins, perhaps Sef1 and Yjl103c themselves, might bind to this sequence and repress transcription. To quantify this interaction, I converted the reporter gene used with the MY1H to *GFP,* and measured the level of gene expression by flow cytometry (Figure 4_3A).

As shown in Figure 4_3A, expression of a reporter gene reglated by the Sef1-AD fusion protein in wild-type cells is much higher than expression with the Yjl103c-AD fusion protein. In a *sef1*Δ strain, expression caused by the Sef1-AD fusion is slightly increased, suggesting that endogenous Sef1 may compete for binding on the promoter. Interestingly, in the same *sef1*Δ strain, expression caused by the Yjl103c-AD is further reduced, leading me to suspect, that Sef1 and Yjl103c may act cooperatively such that in the absence of Sef1, the Yjl103c-AD fusion is less likely to bind to the promoter construct.

The converse also holds true, as expression of the reporter gene stimulated by the Sef1-AD fusion protein in the *yjl103c*Δ strain is significantly reduced, while expression induced by the Yjl103c-AD fusion in the same *yjl103c*Δ strain is increased many fold.

Tempting though it may be to assume that Sef1 and Yjl103c function solely together to form a heterodimer, the results with the *sef1*Δ *yjl103c*Δ strain repudiate this hypothesis. Expression of *GFP* is increased to a greater extent in the double deletion strain by both the Sef1-AD and Yjl103c-AD fusions, indicating that the presence of the other protein is not necessary for binding to the promoter. It may be that Sef1 and Yjl103c function as both heterodimers and homodimers, and that in the absence of the other protein, homodimers regulate expression. There are examples of zinc cluster proteins forming both heterodimers and homodimers at the same binding sites. While an Oaf1 homodimer maintains basal levels of target genes, an Oaf1/Pip2 heterodimer is preferred in the upregulation of genes when cells are grown using oleate as a carbon source (103, 104, 193).

To determine if Sef1 and Yjl103c function as activators or repressors, I employed the *lac*Z reporter under transcriptional control of the bacterial *lexA* operator. Expression of Sef1-LexA or Yjl103c-LexA proteins reduced expression of β-galactosidase 2 to 4.5 fold (Figure 4_3B), depending on the experiment. Both Sef1-LexA and Yjl103c-LexA fusion proteins appear to repress transcription by the same amount.

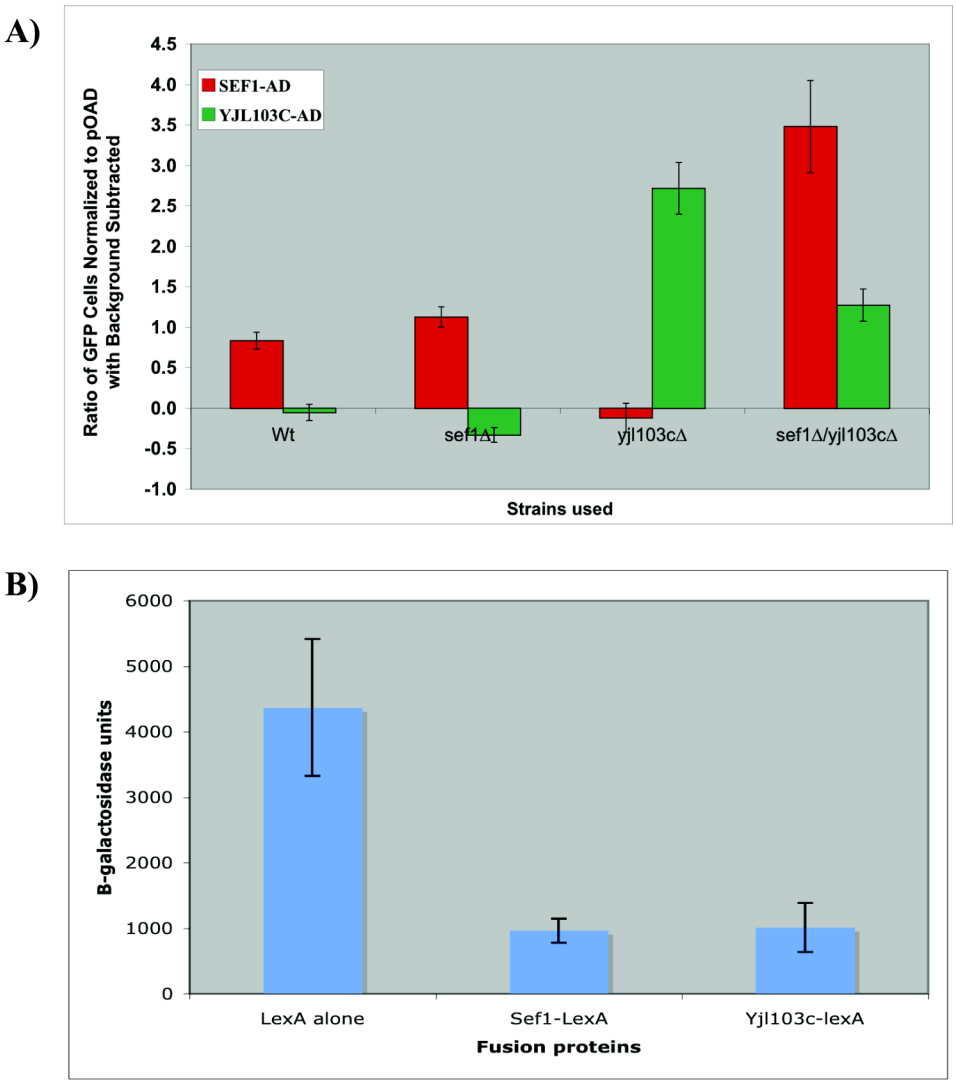


Figure 4_3:Expression of reporter genes

(A) Ratio of cells expressing GFP from a reporter containing Y3A in the promoter in the presence of the Sef1-AD fusion or the Yjl103c fusion proteins. The ratio is normalized to the AD protein and background is subtracted. 10,000 cells are counted per experiment and the average of 3 replicates is used. **(B)** Expression of β-galactosidase from a *lacZ* reporter with lexO binding sites in the presence of LexA alone, a Sef1-lexA fusion protein or a Yjl103c-lexA fusion protein.

Proteins that bind to Sef1 and Yjl103c.

In an attempt to determine if Sef1 and Yjl103c interact, and identify other proteins that they may associate with, I purified two samples of each protein fused to GST and sent them for analysis by matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF). I identified several *S. cerevisiae* proteins predicted from one or more peptides with a significance threshold of $p<0.05$. Yjl103c and Sef1 were identified as one of the top hits in their respective samples.

Yjl103c was associated with 39 different proteins (Table 4_1). 18 were found in both protein samples, while 21 were identified only in one of the two samples. 9 are heat shock and chaperone proteins that have been identified as promiscuous binders (33, 70), 2 are histone proteins and 16 are ribosomal subunits, many of which have been identified in other MALDI-TOF experiments as likely contaminants (33). Of the remaining 11 proteins, 7 are enzymes involved in glycolysis and gluconeogenesis. Yjl103c was previously found to physically interact with Fbp1 (262), another important enzyme in the glycolysis and gluconeogenesis pathways.

Table 4_1:Proteins identified in Maldi-Tof with Yjl103c	
	Adh1, Bmh1, Eno1, Eno2, Fba1, Grx1, Tef1, Tdh1, Ura3
Histone proteins	H2b1, Hhf2
Heat shock/ chaperones	Hsp60, Hsp90, Ssa1, Ssa2, Ssa3, Ssa4, Ssb1, Ssc1, Sse1
Ribosomal proteins	Rpl10, Rpl20b, Rpl31a, Rpl4a, Rpl6b, Rpl8a, Rpp0, Rpp2a, Rps13, Rps16b, Rps17b, Rps2

Sef1 was associated with 49 proteins, 26 of them identified in both protein samples (Table 4_2). After removing the histone (1), chaperones and heat shock proteins (5) and ribosomal subunits (13) from the list, the remaining 30 proteins are enriched in those involved in glycolysis and gluconeogenesis.

Table 4_2:Proteins that bind to Sef1	
	Adh1, Atp2, Bmh1, Cdc19, Cit1, Cpr1, Eno1, Eno2, Fba1, Gal10, Gnd1, Gpm1, Grx1, Hxk1, Hxk2, Ilv5, Mcr1, Mdh1, Pdc1, Pdc6, Pgk1, Por1, Tdh1, Tdh3, Tsa1, Ura3
Histone proteins	Hhf2
Heat shock/ chaperones	Hsp90,Ssa1, Ssa2, Ssa3, Ssb1
Ribosomal proteins	Rpl10, Rpl13b, Rpl20a, Rpl31a, Rpl4a, Rpl6b, Rpp2a, Rps16b, Rps17b, Rps2, Rps22a, Rps5, Rps7a

Yjl103c was not identified in the Sef1 samples, nor vise-versa, consistent with negative results from our pull-down experiments using the two proteins as bait and prey (data not shown). However, it is interesting to note that both proteins associated with the same 9 proteins, most of which are involved in glycolysis and gluconeogenesis (Supplemental Table 6) (*Saccharomyces* Genome Database; http://www.yeastgenome.org).

Gene expression profiling of Sef1.

To identify targets of Sef1 and to look for its interaction with Yjl103c to gain clues of their function, I compared the gene-expression profile of a wild-type strain to those that overexpress or are missing *SEF1*. 135 genes were differentially expressed between the wild-type strain and the *sef1*Δ strain. About half of these were upregulated; about half were down regulated. Genes that were upregulated are enriched in proteins involved in phosphate metabolism ($P= 1.3 X10^{-4}$). 1291 genes were differentially expressed between the wild-type and the *SEF1*-overexpressing strains. 59 of these were upregulated by *SEF1* overexpression, and were enriched in genes involved in protein folding ($P= 4.92 X10^{-6}$). No other significant functional

categories were identified. The paucity of information on *SEF1* in the recent spate of large-scale experiments underscores its elusive nature.

Two genes were highly upregulated in both *yjl103c*Δ and *sef1*Δ strains compared to wild-type. *PHO84*(26), is upregulated 16 fold in the *yjl103c*Δ strain and 12 fold in the *sef1*Δ strain, and *SPL2* (51), is upregulated 15 fold in the *yjl103c*Δ strain, 13 fold in the *sef1*Δ strain. These increases in expression levels were confirmed by RTPCR and by the use of *lacZ* reporters driven by the *PHO84* and *SPL2* promoters (Figure 4_4). The promoters of the genes encoding these proteins do not contain $CGGN_8CGG$, suggesting that loss of repression in both deletion mutants may be an indirect effect.

In response to conditions of phosphate starvation, cells activate the PHO pathway, triggering two feedback elements: a negative feedback loop consisting of *PHO84* induction which helps to bring phosphate into the cell and inactivate the PHO pathway, and a positive feedback loop consisting of up-regulation of *SPL2*, which tends to reduce phosphate uptake, leading to further pathway activation (259) (Figure 4_4E). Phenotypic studies were done on the *yjl103c*Δ, *sef1*Δ and *yjl103c*Δ*sef1*Δ strains to see if growth was affected by different concentrations of inorganic phosphate in the media. All three strains grew similar to the wild-type strain.

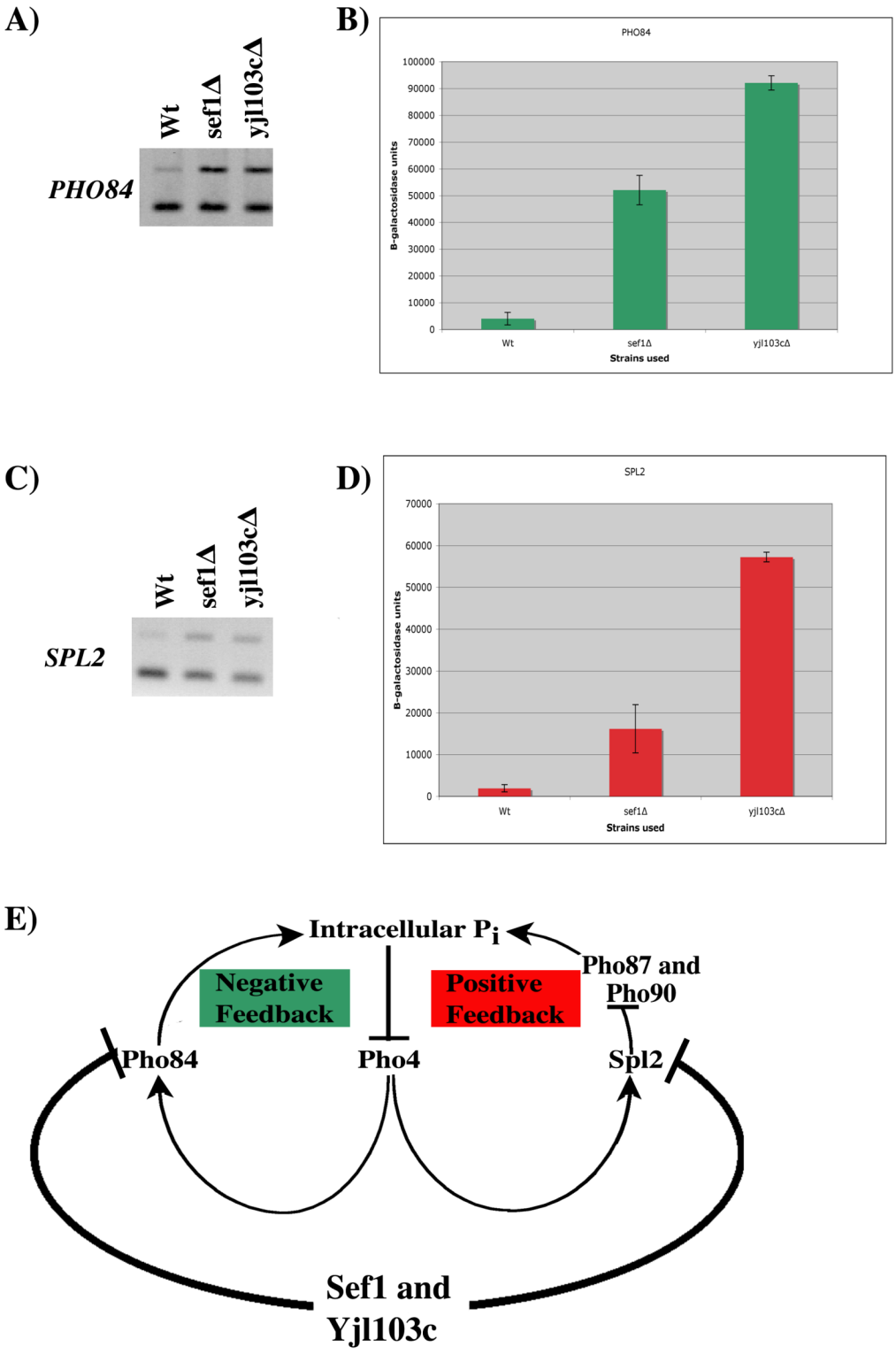


Figure 4_4: Testing expression of *PHO84* and *SPL2* in *yjl103cΔ* and *sef1Δ*.

Figure 4_4: Testing expression of *PHO84* and *SPL2* in *yjl103c*Δ and *sef1*Δ.

(A and C) RTPCR of the **(A)** *PHO84* and **(C)** *SPL2* transcripts in wild-type and deletion strains. **(B and D)** Expression of the *lacZ* reporter driven by the **(B)** *PHO84* and **(D)** *SPL2* promoters in wild-type and deletion strains. **(E)** Model of positive and negative feedback loops in the PHO pathway. Pho84 and Spl2 are the main regulator proteins that cause response to internal inorganic phosphate levels. The bimodal nature of this system allows for sensitive regulation. Sef1 and Yjl103c appear to repress both *PHO84* and *SPL2* expression as expression of both are increased in the *sef1Δ* and *yjl103cΔ* strains. Figure adapted from (180)

Sef1 and Yjl103c have similar but distinct regulatory functions.

I identified 19 target genes of Yjl103c by ChIP (Chapter 3, Figure 3_5). The promoters of these genes all contain $CGGN_8CGG$. To determine if Sef1 and Yjl103c regulate expression of these genes, I used the *lacZ* reporter fused to the promoters of *SDS23, HAP4* and *ADR1*. I also made mutant promoters lacking the $CGGN_8CGG$ binding site. These 3 promoters were also found to bind Yjl103c using the calling card assay (Wang, H, personal communication), giving us more confidence that they are true targets of Yjl103c.

SDS23 expression appears to be regulated by Yjl103c but not by Sef1 as expression in the *sef1*Δ strain is similar to the wild-type strain while expression in the *yjl103c*Δ strain is increased (Figure 4_5A). Regulation by Yjl103c is through the $CGGN_8CGG$ binding site because removing this binding site abolishes the difference in expression due to the deletion strain. Sds23 is a relatively uncharacterized protein

involved in cyclosome regulation (39). It contains a CSRE (carbon source-responsive element) in its promoter, and is upregulated in the presence of a gluconeogenic carbon source (39).

In contrast, *ADR1* and *HAP1* expression appear to be regulated by both Sef1 and Yjl103c (Figure 4_5B and 4_5C): expression in each single mutant is upregulated compared to a wild-type strain. Expression of *ADR1* in the double mutant does not seem very different from that of the single mutants (Figure 4_5B), while expression of *HAP4* in the double mutant is increased compared to the single mutants, suggesting an additive effect of the two proteins (Figure 4_5C) on *HAP4* expression. This regulation by Sef1 and Yjl103c is due to the binding of the proteins to the $CGGN_8CGG$ binding site, as removal of the binding site removes any expression difference between the wild-type and mutant strains.

Adr1 is a transcription factor involved in regulating genes for the utilization of ethanol, glycerol and lactate, coordinating the biochemical pathways that generate acetyl-CoA and NADH from non-fermentable substrates (209, 261). Hap4, another transcription factor, controls the TCA cycle and related pathways (17, 27, 52).

Both *SEF1* and *YJL103C* appear to be regulated by Hap4 as the promoters of both genes contain CCAAT, the consensus binding site for Hap4 (52), suggesting a feedback loop (Figure 4_5D). *YJL103c* expression is significantly changed by *HAP4* deletion in two studies (27, 178) and is strongly correlated with expression of genes involved in oxidative phosphorylation ($P < 10^{-23}$) (41), which are regulated by Hap4.

The Yjl103c promoter contains a $CGGN_8CGG$ sequence, and binding of Yjl103c to its own promoter was confirmed by ChIP (Chapter 3, Figure 3_5).

YJL103C expression is upregulated 5 fold in the *sef1Δ* strain, suggesting that Sef1 also regulates this expression. Sef1, on the other hand, does not have the $CGGN_8CGG$ sequence in its promoter, and shows no difference in expression in the *yjl103cΔ* strain.

The evidence suggests that Yjl103c and Sef1 have similar roles in regulating energy utilization, because they regulate the same genes, *HAP4* and *ADR1*, both important regulators themselves in carbon metabolism. Yet they also have distinct functions as Yjl103c and not Sef1 regulates *SDS23*. They are also regulated differently as Yjl103c forms a negative feedback loop with itself, and appears to be regulated by Sef1, while Sef1 does not appear to be regulated by Yjl103c nor itself (Figure 5_4D).

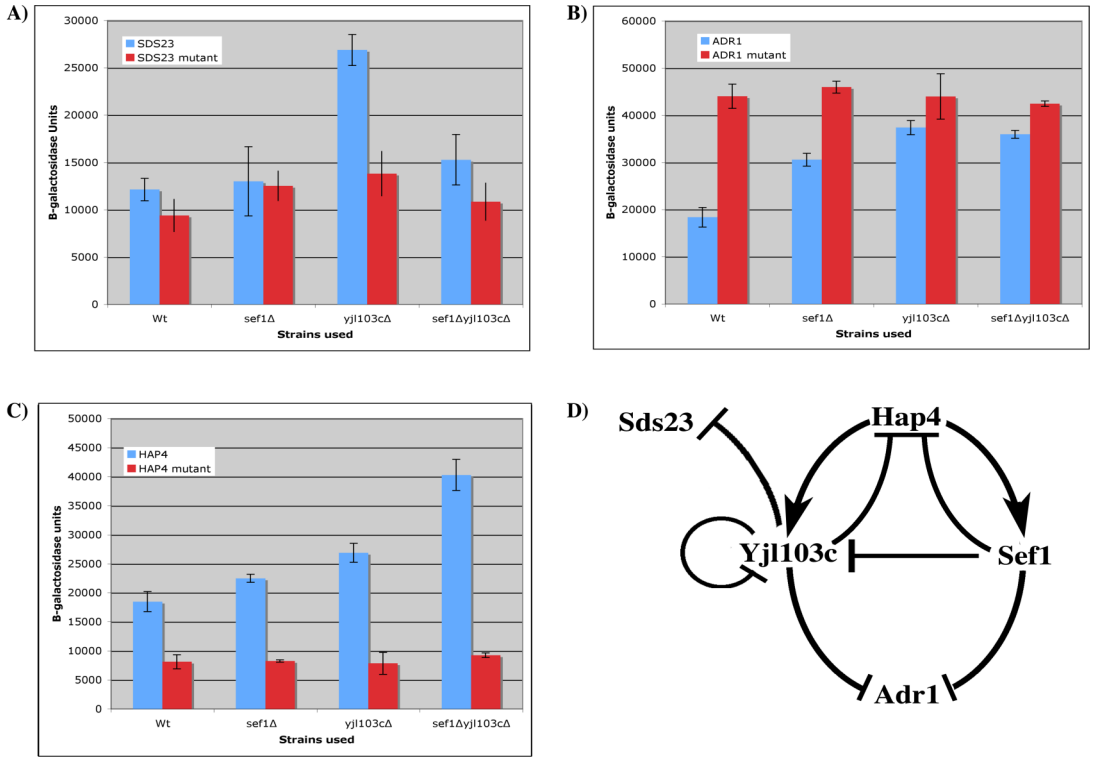


Figure 4_5:
(A), (B), (C): Expression of the *lacZ* reporter driven by the **(A)** *SDS23* and **(B)** *ADR1* and **(C)** *HAP4* promoters in wild-type and deletion strains. Blue bars indicate the wild-type promoter and the red bars indicate the promoters with the $CGGN_8CGG$ binding sites removed. (D) Network of regulation among Adr1, Hap4, Sds23, Sef1 and Yjl103c. *Yjl103c* represses *SDS23* expression. Hap4 forms a feedback loop with both Yjl103c and Sef1. Evidence suggests that Yjl103c and Sef1 may interact to regulate *HAP4* expression. Yjl103c and Sef1 both repress *ADR1* expression, but there is no evidence to suggest that the two proteins interact. Yjl103c and Sef1 both repress *YJL103C* expression, but it does not appear that *Yjl103c* or *Sef1* regulates *SEF1* expression.

Discussion

S. cerevisiae prefers glucose over other carbon sources as it can directly enter the glycolytic pathway. However, it is capable of using alternative energy sources such as galactose, maltose, ethanol and glycerol. The enzymes needed for a specific pathway are usually produced only when required, and their regulation is mainly at the transcriptional level. A shift from fermentation to the nonfermentation mode of growth is characterized by massive changes in expression of genes involved in many different processes such as carbon metabolism, protein synthesis and carbohydrate storage (45). Expression of genes for gluconeogenesis and many respiratory genes, including genes involved in oxidative phosphorylation are coregulated (225).

I found that Sef1 and Yjl103c bind to the sequence $CGGN_8CGG$ *in vitro* and *in vivo*. Sef1 and Yjl103c appear to function as repressors, as their lexA fusion proteins repress expression from reporter genes, and, in the absence of Sef1 and Yjl103c, expression of their target genes is upregulated. Sef1 and Yjl103c regulate genes that contain the sequence $CGGN_8CGG$ in their promoters. Removal of this sequence abolishes regulation by Sef1 and Yjl103c.

Sef1 and Yjl103c interact with several genes involved in glycolysis and gluconeogenesis. Yjl103c binds to the *GFA1 (YKL104C*) and *PYK1 (*YAL038W) promoters (Chapter 3_5), which contain the $CGGN_8CGG$ sequence. Gfa1 catalyzes conversion of fructose-6 phosphate, an intermediate of glycolysis, to glucosamine-6-P and glutamate (246). Pyk1 functions in glycolysis to catalyze conversion of phosphoenolpyruvate to pyruvate, the input for aerobic (TCA cycle) or anaerobic (glucose fermentation) respiration, and exerts significant control over the rate and

direction of carbon flux in yeast (246). Yjl103c also regulates expression of *FBP1* and *PCK1,* two gluconeogenic genes (225). This suggests that Sef1 and Yjl103c play a role in regulating expression of genes involved in glycolysis and gluconeogenesis.

SEF1 and *YJL103C* regulate, and are regulated, by *HAP4*, a component of the Hap2/3/4/5 transcription factor complex that controls the expression of genes coding for all the respiratory chain complexes and the enzymes of the TCA cycle (27). Mutations in *YJL103C* are synthetically lethal with mutations in the gene encoding Hap5 (33), suggesting that the interplay between Yjl103c and the Hap2/3/4/5 complex may be important for cell viability.

SEF1 and *YJL103C* also regulate *ADR1* expression, which encodes another central regulator of multiple metabolitic pathways (261) that is responsible for part of the altered transcriptional program that accompanies depletion of glucose. Thus, it appears that Sef1 and Yjl103c both function as transcriptional regulators of several metabolitic pathways.

PHO84 and *SPL2*, regulators in PHO pathway, are highly upregulated in *sef1Δ* and *yjl103cΔ* strains. The PHO pathway is involved in regulating intracellular levels of inorganic phosphate. While individual nutrient pathways have been studied extensively, little is known about the converging effector branches that orchestrate the dynamic responses to nutritional cues (217). There is evidence that shows inorganic phosphate acts in concert with glucose as the nutrient signal for activation of the cAMP–protein kinase A (PKA) pathway (65). The PKA pathway in *S. cerevisiae* plays a major role in the control of metabolism, stress resistance and proliferation, in particular in connection with available nutrients (220). Yjl103c and Sef1 may

coordinate the sensing of various nutrient sources, consistent with their role in regulating both *ADR1* and *HAP4* expression, two major regulators in various carbon metabolism pathways.

Sef1 and Yjl103c have similar but distinct roles in regulating metabolism. *SEF1* and *YJL103C* are upregulated greater than 12-fold in the presence of glycerol or ethanol compared with glucose (191), but they do not seem to be co-regulated in other conditions. In a strain overexpressing *YJL103C,* more than 500 genes were differentially expressed including genes enriched in carbon compound and carbohydrate metabolism, as well as genes involved in stress response (Chapter 3). A larger number of genes are differentially expressed in a *sef1Δ* strain and in the strain overexpressing *SEF1*, yet there was no significant enrichment in functional classes of genes besides a slight enrichment in genes involved in phosphate metabolism.

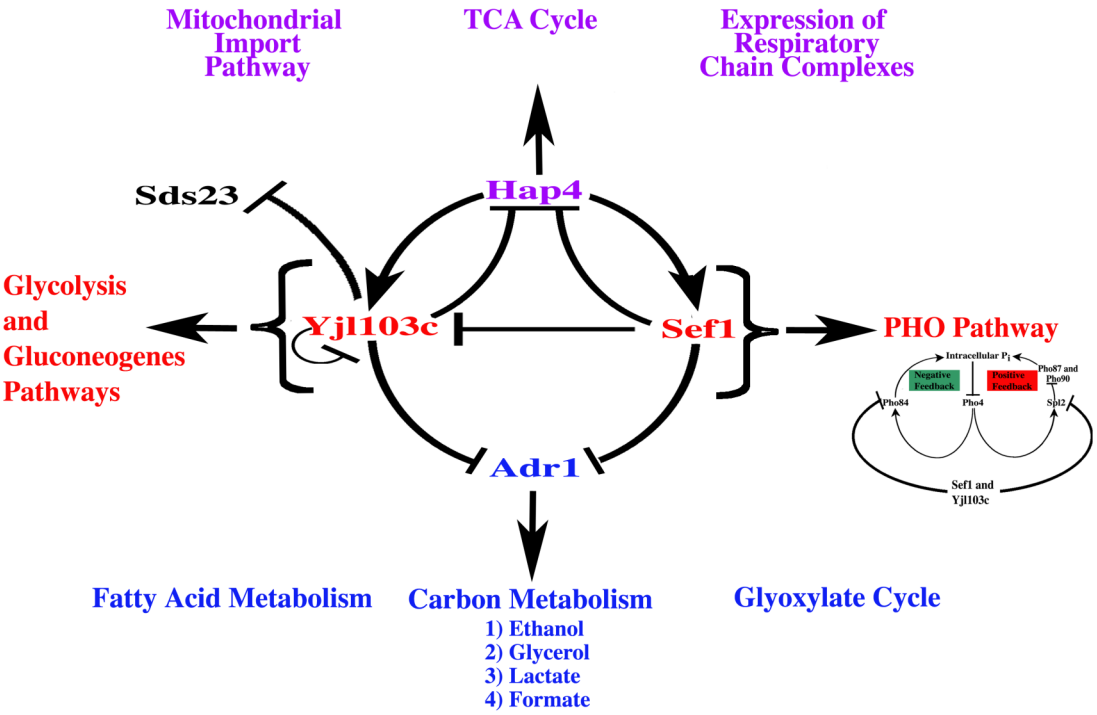


Figure 4_6: Regulatory network involving Yjl103c, Sef1, Hap4 and Adr1. Transcription factors and the processes that they regulate are in the same color. Adr1 regulates fatty acid metabolism, various non-fermentable carbon utilization and the glyoxylate cycle. Hap4 regulates genes involved in the mitochondrial import pathway, the TCA cycle and the genes encoding proteins involved in the respiratory chain complexes in the mitochondria. Together, Adr1 and Hap4 control most aspects of the non-fermentable carbon metabolic pathways. Yjl103c and Sef1 regulate genes involved in glycolysis and gluconeogenesis in addition to regulating *ADR1* and *HAP4* expression. Hap4 regulates *SEF1* and *YJL103C* expression, forming a feedback loop. Sef1 and Yjl1103c regulate *YJL103C* expression but *SEF1* expression does not appear to be regulated by either protein. Both Sef1 and Yjl103c regulate the

regulators of the PHO pathway, and may aid in integrating the different nutrient pathways (65, 217).

A model of how *SEF1* and*YJL103C* interact in a regulatory network governing utilization of nonfermentable carbon sources is shown above (Figure 4_6). The use of two regulators in the same pathway is illustrated by Mth1 and Std1, which both regulate glucose metabolism through their interaction with Rgt1, with Mth1 acting as the primary regulator and Std1 serving to buffer the response to glucose (196). The positive and negative feedback loop of the PHO pathway using Spl2 and Pho84 is another example of how two similar regulatory loops provide sensitive response to environmental and internal conditions. Sef1 and Yjl103c play similar but distinct roles in the regulation of non-fermentable carbon metabolism pathways through their regulation of *ADR1* and *HAP4* expression and through their regulation of genes involved in gluconeogenesis and glycolysis.

Materials and Methods

Materials and Methods

Strains and Plasmids

S. cerevisiae strains used in this study are listed in Supplemental Table 7.

All of the yeast cultivation was done at 30 °C, in flasks, shaken at 325 rpm unless otherwise specified. Synthetic complete medium (SC) lacking the appropriate amino acid (2 g/liter synthetic dropout mix (US Biological), 1.7 g/liter yeast nitrogen base, 5 g/liter ammonium sulfate) and supplemented with the indicated carbon source was used in all cultivations in which prototrophic selection was necessary. All of the yeast transformations were performed according to standard methods (64). All of the plasmids were constructed via gap repair (179) by PCR amplification of sequence to be inserted, flanked by 20–25 bp of homology to the recipient linearized plasmid. Plasmids in positively selected clones were recovered from yeast and transformed into Escherichia coli GC10 Thunderbolt (GeneChoice, Inc.) electrocompetent cells for amplification and DNA sequencing.

Yeast-1 hybrid

We used PJ69-4a to carry the AD fusions. These strains were mated to *HIS3* reporter strains derived from PJ69-4α (95). *HIS3* reporter plasmid pBM4429 was based on pRS316 (backbone CEN plasmid with *URA3*) (208). Three overlapping PCR products were inserted into the backbone by gap-repair to produce a *MEL1* minimal promoter interrupted by *TRP1*, flanked by Spe1 and Xho1 sites. The resulting plasmid was cut with Spe1 and Xho1 and gel-purified for gap repair with the

double-stranded motif. The *GFP* reporter pBM4796 was created by replacing *HIS3* in pBM4429 with *GFP*.

Probe preparation for protein microarrays

Probes were made by a fill-in reaction with *Taq* DNA polymerase using a universal oligonucleotide labeled with Cy3 or Cy5 (for protein chips) or biotin (for EMSA). Probes were purified, concentrated, and quantified by acrylamide gel electrophoresis and a NanoDrop apparatus (NanoDrop Technologies).

Transcription factors tagged on their N-termini with GST-His_6 (266) were over-expressed in yeast cells and purified from 100 ml cultures grown to mid-log phase in 1% yeast extract and 2% peptone and induced for 5 hrs with 2% galactose. Proteins were purified from cell extracts in 96 deep well plates using glutathione beads (GE Healthcare) as previously described (266).

Protein Microarrays

The GST-tagged transcription factors were arrayed into 384 microwell plates, and printed on FAST slides (8 pads, 16 pads or single pad slides; Schleicher & Schuell) in duplicate, triplicate or quadruplicate. In pilot experiments, nickel coated (XENOSLIDE N; Xenopore, Inc.) and aldehyde coated slides (SMAI; Telechem, Inc.) were also tested. We chose the FAST slides because of their higher capacity for protein.

The protein microarrays were probed (in duplicate) with labeled oligonucleotides using the following protocol. Printed slides were blocked for 1 hour

with 3% BSA in hybridization buffer (25 mM HEPES pH 8.0, 50 mM KCl, 0.5% Triton-X100, 2 mM $MgCl_2$, 1 mM PMSF, 3 mM DTT, and protease inhibitors (Complete, Roche)), then probed for 90 mins. with 40 nM fluorescently labeled double stranded DNA oligonucleotides (see Supplemental Table 4 for list of oligonucleotides) in hybridization buffer at 4^{o}C, washed 3 times in cold hybridization buffer and air-dried. The slides were scanned with an Axon GenePix 4000 scanner. Proteins whose signal was reproducibly above background levels ($n \geq 2$ slides) and specific for the wild-type probes, were classified as putative targets and tested further as described below. The conditions for binding were chosen based on extensive experimentation with a wide variety of conditions for binding of probes to Rap1 and Zap1, including different buffers (Tris-HCl and Tris-Borate and HEPES) at a variety of concentrations (25mM to 150 mM) at pHs between 7.0 and 8.0 with different salts (KCl and NaCl) at several different concentrations (25 mM to 150 mM) and different temperatures. Non-specific carrier DNA (salmon sperm DNA and poly dI-dC) was omitted from the final protocol because it increased the background signal of labeled probe to the nitrocellulose surface. Glycerol in the binding buffer higher than 20% smeared the slides; 10% glycerol seemed optimal. Neither Triton-X100 nor Tween-20 detergents (0.1% to 10%) had an observable effect on binding of probe to the arrays. Binding of probes to the array increased with probe concentration to about 50 nM, after which increased background binding to the nitrocellulose surface of the slides was observed.

Electrophoretic Mobility Shift Assays

Binding reactions were carried out according to manufacturer's recommendations (Pierce - Light shift Chemiluminescent Kit), in 20μl (50 nM KCl, 25 mM HEPES, pH 8, 10% glycerol, 0.1% Triton-X). Probe concentrations varied from 60 pM to 600 pM. Protein concentrations varied from 0.7 μM to 8.5 μM. The reactions were incubated for 20 mins. at room temperature followed by 10 mins. on ice before 5 μl of 5% Ficoll loading dye was added and loaded onto 8 cm by 7 cm 8% acrylamide gels pre-run at 100V for one hour. The gels were run at 4^{O}C, 100 V until the bromophenol blue dye had migrated two thirds of the way down the gel. Nucleic acids were transferred to nylon membranes and visualized according to the manufacturer's recommendations.

DNA Microarrays

DNA microarrays were printed with 6,388 oligonucleotides manufactured by Qiagen-Operon that represent virtually all *S. cerevisiae* open reading frames. The oligonucleotides were resuspended to a concentration of 40 μM in 3x SSC with 0.75 M betaine and were printed in duplicate on Epoxy slides (MWG Biotech).

RNA preparation

Cells were grown to log phase in YP 2% raffinose medium, and induced with galactose for 5hrs. RNA extraction, labeling and hybridization was done as described (47).

Data analysis

The scanned the array images were analyzed using the default settings in GenePix Pro 4.0*. For each spot on the array, the median of the pixel-by-pixel ratios of the two channel intensities (with median background intensity subtracted) was calculated, and the two-step mixed model analysis of variance (ANOVA) was applied to the log transformed values (254) and used to normalize the expression differences between spots that were due to factors we were not interested in:

$$\log_2(Y_{ijkm}) = \mu + G_i + T_j + A_{k(ij)} + GT_{ij} + \varphi_{(ijk)m}$$

Where Y_{ijkm} is the median of ratios for each spot, G_i is the average genotype effect (over-expressed or delta strain); T_j is the average treatment effect (wild-type or modified strain),

GT_{ij} is the average genotype x treatment interaction effect (wild-type over-expressed strain, wild-type delta strain, modified over-expressed strain, modified delta-strain), $A_{k(ij)}$ is the average array effect, which is nested within the genotype by treatment interaction effect, $\varphi_{(ijk)m}$ is the residual.

A second ANOVA model was applied to each gene separately using the residual $\varphi_{(ijk)m}$ from each spot as a response variable:

$$\varphi_{gijkm} = \gamma_g + \gamma G_{gi} + \gamma T_{gj} + \gamma GT_{gij} + \gamma A_{gk(ij)} + \varepsilon_{(gijk)m}$$

Where φ_{gijkm} is the residual from the first ANOVA model for each spot, γ_g is the average gene expression for each gene *g*, G_{gi} is the gene expression due to genotype *i*, T_{gj} is the gene expression due to treatment *j*, GT_{gij} is the gene expression

due to genotype *i* interacting with treatment *j*, $A_{gk(ij)}$ is the gene expression due to array effect, $\varepsilon_{(gijk)m}$ is the residual.

Genes that showed differential expression between wild-type (GT_{10}) and over-expressed (GT_{11}) strains were selected based on the criteria: $\gamma GT_{10} - \gamma GT_{11} \neq 0$ at α=0.05. Genes that showed differential expression between wild-type (GT_{00}) and delta (GT_{01}) strains were selected based on similar criteria: $\gamma GT_{00} - \gamma GT_{01} \neq 0$ at α=0.05. To select genes that show differential expression between the over-expressed and delta strains, several filters were applied. First, genes that satisfied $\gamma GT_{01} - \gamma GT_{11} \neq 0$ at α=0.05 were kept. Next, genes that satisfied $\gamma GT_{00} - \gamma GT_{10} \neq 0$ at α=0.05 were filtered out. Last, we filtered out genes that did not show any significance for the genotype by treatment interaction effect.

Chromatin Immunoprecipitation (ChIP)

Yjl103 was expressed from its own promoter and tagged at its C-terminus with 13 copies of the myc epitope by integrating into the chromosome 13Myc-KanMX, as previously described (129).The strain expressing Yjl103 tagged with the myc epitope, and the corresponding wild-type strain were inoculated at an $O.D._{600}$ of 0.2 and grown in YPD medium overnight and reinoculated into fresh medium at an $O.D._{600}$ of 0.2. The strains were grown for 4 hrs at 30^{o}C with shaking before heated medium was added to bring the temperature of the cultures to 37^{o}C. After 20 mins. of shaking at 37 ^{o}C cells were fixed by addition of formaldehyde to a final concentration of 1%. Proteins were precipitated using 9E10 anti-c-myc antibody (Santa Cruz

Biotech), and the associated DNA was liberated, purified, amplified and labeled with Cy3 and Cy5 fluorophores and used to probe a DNA microarray of intergenic regions of the yeast genome, as previously described (93). Twenty-two gene promoters that showed an increased hybridization signal relative to the signal obtained with probe prepared from the immunoprecipitate of a strain without a myc-tagged protein, and whose genes were differentially expressed in strains missing or over expressing Yjl103 (determined by a gene expression profiling experiment as described above) were chosen for further analysis by conventional ChIP, preformed as described above except that 40 ng of the liberated DNA was amplified (30 cycles at 95 ^{o}C for 1.5 min, 57 ^{o}C for 2 min, and 72 ^{o}C for 3 min, with a final extension at 72 ^{o}C for 10 min) in a 50 μl reaction with Mango Taq (BioLine). Two sets of primers were used in each reaction. Primers to query promoters were designed to generate a product of about 500 base pairs and were added to a final concentration of 2nM; primers to the *GAL4* promoter generated a product of 150 base pairs and were added to a final concentration of 0.4 nM.

MALDI-TOF

Strains expressing Sef1 and Gsm1 tagged at their C-terminus with GST on plasmids were inoculated at inoculated at an O.D.$_{600}$ of 0.02 and grown in –ura raffinose medium for 12 hrs at 30^{o}C with shaking before induction with a final concentration of 2% galactose for 6 hours. The fusion proteins were purified by pull-down with glutathione beads. Two independent samples of each protein were sent to

the Proteomics Core facility at the Siteman Cancer Center for analysis by MALDI-TOF mass spectrometry.

β-Galactosidase Assay

Enzyme activity was measured in aliquots of cell culture lysed in Y-PER (Pierce). The enzymatic activity in each lysate was determined by monitoring the increase in fluorescence caused by the liberation of 4-methylumbelliferone through cleavage of the glycosidic bond by the β-galactosidase enzyme. This assay is based on the Betafluor β-galactosidase assay kit (Novagen). Briefly, the A_{600} of 100 µl of cell culture was measured on a Bio-Tek Synergy HT and immediately following the reading, 50 µl of culture was added to 25 µl of Y-PER reagent to lyse the cells present. To this lysate, 75 µl of reaction mix was added. The reaction mix contains Z buffer, pH 7.0 (60 mm Na_2HPO_4, 40 mm NaH_2PO_4, 10 mm KCl, 1 mm $MgSO_4$, 1 mm dithiothreitol) (150), and the fluorogenic substrate 4-methylumbelliferyl β-d-galactopyranoside (Sigma) dissolved in dimethyl sulfoxide to 20 mg/ml and present at a final concentration of 1 mg/ml. The progress of the reaction was monitored at 30 °C in a Bio-Tek Synergy HT plate reader (excitation, 360 ± 40 nm; emission, 460 ± 40 nm; sensitivity, 50), with readings taken every 2 min after 3 s of shaking. The β-galactosidase units reported were calculated according to the following formula.

$$\text{β-gal units} = \frac{\text{Initial velocity (AFU min}^{-1}\text{)}}{A_{600} \times V_{culture}(\mu l)}$$

RT-PCR

Cells grown to log phase in synthetic complete medium were diluted to an OD_{600} of 0.2 to 0.3 in synthetic complete medium in glucose and grown for 2.5 h at 30°C. Cells were heat-shocked for 15 min by addition of hot media to bring the temperature up to 37°C. Cell pellets were snap-frozen in liquid nitrogen and stored at −80°C. For total RNA purification, cells were resuspended in 400 µl of Tris-EDTA plus 0.5% sodium dodecyl sulfate and 500 µl of hot acid-phenol (65°C) and kept at 65°C for 1 h, with vortexing for 10 s every 10 min. The RNA was phenol-chloroform extracted and ethanol precipitated. A total of 5 µg of RNA was treated with DNase (Ambion DNA-free, catalogue no. 1907), and 500 ng was used in a 20-µl reverse transcription (RT) reaction (Superscript II; Invitrogen). One microliter of cDNA was used in each 20-µl PCR.

Supplemental Tables

Supplemental Table 1. 169 TF-AD used in one-hybrid that identified Stp2

A2		**MATALPHA2**		**STP1**		**YLR013W**
ABF1		**MBP1**		**STP2**		**YLR074C**
ADR1		**MCM1**		**STP3**		**YLR266C**
ALPHA2		**MET28**		**SUM1**		**YML076C**
ARG81		**MET31**		**SWI3**		**YMR136W**
ARGR1		**MET32**		**SWI5**		**YNL227C**
ARR1		**MGA1**		**TEC1**		**YNR063W**
ASH1		**MGA2**		**TFC2**		**YOL089C**
AZF1		**MIG1**		**THI2**		**YOR172W**
BUR6		**MIG2**		**TUP1**		**YOR380W**
CAD1		**MOT3**		**TYE7**		**YPL133C**
CBF1		**MSN2**		**UGA3**		**YPR008W**
CEF1		**NCB2**		**UME6**		**YPR013C**
CEP3		**NHP10**		**YAP3**		**YPR015C**
CIN5		**PEP7**		**YAP5**		**YPR196W**
CRZ1		**PHD1**		**YAP7**		**YOX1**
CSE2		**PHO2**		**YBL010C**		**YRR1**
CTH1		**PHO4**		**YBL054W**		**ZAP1**
CTH2		**PIP2**		**YBR033W**		**ZMS1**
CUP2		**PPR1**		**YBR239C**		
CUP9		**PRP11**		**YBR267W**		
DAL80		**PRP9**		**YCR106W**		
DAL81		**RAD18**		**YDL098C**		
DOT6		**REB1**		**YDR026C**		
FKH1		**RFX1**		**YDR049W**		
FKH2		**RGM1**		**YDR112W**		
FZF1		**RGT1**		**YDR213W**		
GAT1		**RIM101**		**YDR303C**		
GCN4		**RLM1**		**YDR421W**		
GIS1		**RME1**		**YDR451C**		
GZF3		**ROX1**		**YER028C**		
HAC1		**RPN4**		**YER045C**		
HAP3		**RSC8**		**YER130C**		
HAP4		**RTG1**		**YER169W**		
HAP5		**RTG3**		**YER184C**		
HCM1		**RTS2**		**YFL044C**		
HMO1		**SAS2**		**YFL052W**		

HMS1		**SAS3**		**YGL096W**		
HMS2		**SEF1**		**YGR002C**		
HSF1		**SFL1**		**YGR067C**		
INO2		**SFP1**		**YHR207C**		
INO4		**SIP4**		**YIL036W**		
IXR1		**SKN7**		**YIL130W**		
LEE1		**SKO1**		**YIR013C**		
LEU3		**SMP1**		**YJL206C**		
LYS14		**SOK2**		**YJL206CA**		
MAC1		**SPT23**		**YJR119C**		
MAL13		**SSN21**		**YKL222C**		
MAL33		**SSN22**		**YKR064W**		
MATA1		**STB4**		**YLL054C**		
		STB5				

Supplemental Table 2. 269 TF-AD used in one-hybrid that identified Sef1

A2	**PIP2**	**YCR047C**	**YKR099W**
ABF1	**PPR1**	**YCR096C**	**YLL054C**
ADR1	**PRP11**	**YCR106W**	**YLR013W**
ALPHA2	**PRP9**	**YDL020C**	**YLR039C**
ARG81	**RAD18**	**YDL023C**	**YLR074C**
ARGR1	**RDS2**	**YDL043C**	**YLR098C**
ARR1	**REB1**	**YDL098C**	**YLR131C**
ASH1	**RFX1**	**YDL160C**	**YLR182W**
AZF1	**RGM1**	**YDL197C**	**YLR228C**
BUR6	**RGT1**	**YDL231C**	**YLR256W**
CAD1	**RIM101**	**YDR009W**	**YLR266C**
CBF1	**RLM1**	**YDR026C**	**YLR403W**
CEF1	**RME1**	**YDR049W**	**YLR418C**
CEP3	**ROX1**	**YDR112W**	**YML007W**
CIN5	**RPN4**	**YDR157W**	**YML010W**
CRZ1	**RSC8**	**YDR213W**	**YML051W**
CSE2	**RTG1**	**YDR257C**	**YML076C**
CTH1	**RTG3**	**YDR303C**	**YML081W**
CTH2	**RTS2**	**YDR359C**	**YML081W**
CUP2	**SAS2**	**YDR360W**	**YMR039C**
CUP9	**SAS3**	**YDR421W**	**YMR136W**
DAL80	**SEF1**	**YDR451C**	**YMR176W**
DAL81	**SFL1**	**YDR520C**	**YMR179W**
DOT6	**SIP4**	**YER028C**	**YMR228W**
FKH1	**SKN7**	**YER040W**	**YMR280C**
FKH2	**SKO1**	**YER045C**	**YMR291W**
FZF1	**SMP1**	**YER111C**	**YNL103W**
GAT1	**SOK2**	**YER130C**	**YNL139C**
GCN4	**SPT23**	**YER164W**	**YNL140C**
GIS1	**SSN21**	**YER169W**	**YNL216W**
GZF3	**SSN22**	**YER184C**	**YNL222W**
HAC1	**STB4**	**YFL044C**	**YNL227C**
HAP3	**STB5**	**YFL052W**	**YNL251C**
HAP4	**STP1**	**YFL063W**	**YNL314W**
HAP5	**STP2**	**YFR017C**	**YNR063W**
HCM1	**STP3**	**YGL013C**	**YOL004W**
HMO1	**SUM1**	**YGL096W**	**YOL006C**
HMS1	**SWI3**	**YGL162W**	**YOL012C**
HMS2	**SWI5**	**YGL208W**	**YOL089C**
HSF1	**TEC1**	**YGL244W**	**YOR038C**
INO2	**TFC2**	**YGR002C**	**YOR172W**

INO4
IXR1
LEE1
LEU3
LYS14
MAC1
MAL13
MAL33
MATA1
MATALPHA2
MBP1
MCM1
MET28
MET31
MET32
MGA1
MGA2
MIG1
MIG2
MOT3
MSN2
NCB2
NHP10
PEP7
PHD1
PHO2
PHO4

THI2
TUP1
TYE7
UGA3
UME6
YAP3
YAP5
YAP7
YAL051W
YBL008W
YBL010C
YBL049W
YBL054W
YBL065W
YBR033W
YBR063C
YBR083C
YBR112C
YBR150C
YBR215W
YBR239C
YBR267W
YBR279W
YCL042W
YCL048W
YCR040W
YCR043C

YGR063C
YGR067C
YGR116W
YGR146C
YGR200C
YGR272C
YHR056C
YHR119W
YHR207C
YIL010W
YIL036W
YIL038C
YIL128W
YIL130W
YIR013C
YIR042C
YJL103C
YJL115W
YJL147C
YJL206C
YJL206CA
YJR119C
YKL015W
YKL062W
YKL222C
YKL223W
YKR064W

YOR229W
YOR337W
YOR376W
YOR380W
YPL016W
YPL021W
YPL082C
YPL128C
YPL129W
YPL133C
YPL230W
YPL254W
YPR008W
YPR009W
YPR013C
YPR015C
YPR022C
YPR072W
YPR104C
YPR196W
YOX1
YRR1
ZAP1
ZMS1

Supplemental Table 3: DNA sequences inserted into the promoter

Y1A	AGTCAGGACGCATTGGCTTTGGTACGTATTGGTACGTATGGATCCCGCTCATTTCTAGTGGTACGTATTGGTACGTATAAGAAGAAAGGAAGACATGT
Y1B	AGTCAGGACGCATTGGCTTTGGTACGTATTGGTACGTATGGATCCCGCTCATTTCTAGTGGTACGTATTGGTACGTATAAGAAGAAAGGAAGACATGT
Y2A	AGTCAGGACGCATTGGCTTTTTCTAGATTTCTAGAGGATCCCGCTCATTTCTAGTTTCTAGATTTCTAGAAAGAAGAAAGGAAGACATGT
Y2B	AGTCAGGACGCATTGGCTTTATCTAGATATCTAGAGGATCCCGCTCATTTCTAGTATCTAGATATCTAGAAAGAAGAAAGGAAGACATGT
Y3A	AGTCAGGACGCATTGGCTTCTACCTCCGAGCTACCTCCGAGGGATCCCGCTCATTTCTAGCTACCTCCGAGCTACCTCCGAGAAGAAGAAAGGAAGACATGT
Y3B	AGTCAGGACGCATTGGCTTTAGCGATAGCGAGGATCCCGCTCATTTCTAGTAGCGATAGCGAAAGAAGAAAGGAAGACATGT
Y4A	AGTCAGGACGCATTGGCTTACCAAGCAAAAACCAAGCAAAAGGATCCCGCTCATTTCTAGACCAAGCAAAAACCAAGCAAAAAGAAGAAAGGAAGACATGT
Y4B	AGTCAGGACGCATTGGCTTACGAACCATAAACGAACCATAAGGATCCCGCTCATTTCTAGACGAACCATAAACGAACCATAAAGAAGAAAGGAAGACATGT
Y5A	AGTCAGGACGCATTGGCTTAACAACAACAACGGATCCCGCTCATTTCTAGAACAACAACAACAAGAAGAAAGGAAGACATGT
Y5B	AGTCAGGACGCATTGGCTTAATGACAATGACGGATCCCGCTCATTTCTAGAATGACAATGACAAGAAGAAAGGAAGACATGT
Y6A	AGTCAGGACGCATTGGCTTAACTTTTCAACTTTTCGGATCCCGCTCATTTCTAGAACTTTTCAACTTTTCAAGAAGAAAGGAAGACATGT
Y6B	AGTCAGGACGCATTGGCTTTTCTTTTGTTCTTTTGGGATCCCGCTCATTTCTAGTTCTTTTGTTCTTTTGAAGAAGAAAGGAAGACATGT
Y7A	AGTCAGGACGCATTGGCTTAAGCCACAAAGCCACAGGATCCCGCTCATTTCTAGAAGCCACAAAGCCACAAAGAAGAAAGGAAGACATGT
Y7B	AGTCAGGACGCATTGGCTTAACTCACAAACTCACAGGATCCCGCTCATTTCTAGAACTCACAAACTCACAAAGAAGAAAGGAAGACATGT
Y8A	AGTCAGGACGCATTGGCTTACATACAACATACAGGATCCCGCTCATTTCTAGACATACAACATACAAAGAAGAAAGGAAGACATGT
Y8B	AGTCAGGACGCATTGGCTTAGTTGCAGTTGCGGATCCCGCTCATTTCTAGAGTTGCAGTTGCAAGAAGAAAGGAAGACATGT
Y9A	AGTCAGGACGCATTGGCTTAGAGAGAGGGATCCCGCTCATTTCTAGAGAGAGAGAAGAAGAAAGGAAGACATGT
Y9B	AGTCAGGACGCATTGGCTTACACTGACACTGGGATCCCGCTCATTTCTAGACACTGACACTGAAGAAGAAAGGAAGACATGT
Y10A	AGTCAGGACGCATTGGCTTAGTGAAAGTGAAGGATCCCGCTCATTTCTAGAGTGAAAGTGAAAAGAAGAAAGGAAGACATGT
Y10B	AGTCAGGACGCATTGGCTTAGTCAAAGTCAAGGATCCCGCTCATTTCTAGAGTCAAAGTCAAAAGAAGAAAGGAAGACATGT
Y11A	AGTCAGGACGCATTGGCTTATATGTATATGTGGATCCCGCTCATTTCTAGATATGTATATGTAAGAAGAAAGGAAGACATGT
Y11B	AGTCAGGACGCATTGGCTTAGATTGAGATTGGGATCCCGCTCATTTCTAGAGATTGAGATTGAAGAAGAAAGGAAGACATGT
Y12A	AGTCAGGACGCATTGGCTTATGCGATGCGGGATCCCGCTCATTTCTAGATGCGATGCGAAGAAGAAAGGAAGACATGT
Y12B	AGTCAGGACGCATTGGCTTACGCGAAGACGCGAAGGGATCCCGCTCATTTCTAGACGCGAAGACGCGAAGAAGAAGAAAGGAAGACATGT
Y13A	AGTCAGGACGCATTGGCTTCAAACAAACAAACAAAGGATCCCGCTCATTTCTAGCAAACAAACAAACAAAAAGAAGAAAGGAAGACATGT
Y13B	AGTCAGGACGCATTGGCTTCAAACAAACAAACAAAGGATCCCGCTCATTTCTAGCAAACAAACAAACAAAAAGAAGAAAGGAAGACATGT
Y14A	AGTCAGGACGCATTGGCTTCAAGGGCAAGGGGGATCCCGCTCATTTCTAGCAAGGGCAAGGGAAGAAGAAAGGAAGACATGT
Y14B	AGTCAGGACGCATTGGCTTCAAGTGCAAGTGGGATCCCGCTCATTTCTAGCAAGTGCAAGTGAAGAAGAAAGGAAGACATGT
Y15A	AGTCAGGACGCATTGGCTTCACCACCACCACGGATCCCGCTCATTTCTAGCACCACCACCACAAGAAGAAAGGAAGACATGT

Y15B	AGTCAGGACGCATTGGCTTCACTACCACTACGGATCCCGCTCATTTCTAGCACTACCACTACAAGAAGAAAGGAAGACATGT
Y16A	AGTCAGGACGCATTGGCTTAATAACAATAACGGATCCCGCTCATTTCTAGAATAACAATAACAAGAAGAAAGGAAGACATGT
Y16B	AGTCAGGACGCATTGGCTTCGACACCGACACGGATCCCGCTCATTTCTAGCGACACCGACACAAGAAGAAAGGAAGACATGT
Y17A	AGTCAGGACGCATTGGCTTTACACCTACACCGGATCCCGCTCATTTCTAGTACACCTACACCAAGAAGAAAGGAAGACATGT
Y17B	AGTCAGGACGCATTGGCTTCTCACGCTCACGGGATCCCGCTCATTTCTAGCTCACGCTCACGAAGAAGAAAGGAAGACATGT
Y18A	AGTCAGGACGCATTGGCTTCTAAACGACTAAACGAGGATCCCGCTCATTTCTAGCTAAACGACTAAACGAAAGAAGAAAGGAAGACATGT
Y18B	AGTCAGGACGCATTGGCTTCGATACTACCGATACTACGGATCCCGCTCATTTCTAGCGATACTACCGATACTACAAGAAGAAAGGAAGACATGT
Y19A	AGTCAGGACGCATTGGCTTCTGAAAACTGAAAAGGATCCCGCTCATTTCTAGCTGAAAACTGAAAAAAGAAGAAAGGAAGACATGT
Y19B	AGTCAGGACGCATTGGCTTCAGAATACAGAATAGGATCCCGCTCATTTCTAGCAGAATACAGAATAAAGAAGAAAGGAAGACATGT
Y20A	AGTCAGGACGCATTGGCTTGCCAAGGCCAAGGGATCCCGCTCATTTCTAGGCCAAGGCCAAGAAGAAGAAAGGAAGACATGT
Y20B	AGTCAGGACGCATTGGCTTGCTAACGCTAACGGATCCCGCTCATTTCTAGGCTAACGCTAACAAGAAGAAAGGAAGACATGT
Y21A	AGTCAGGACGCATTGGCTTGCGATGCGCGATGCGGATCCCGCTCATTTCTAGGCGATGCGCGATGCAAGAAGAAAGGAAGACATGT
Y21B	AGTCAGGACGCATTGGCTTGTGATTCGTGATTCGGATCCCGCTCATTTCTAGGTGATTCGTGATTCAAGAAGAAAGGAAGACATGT
Y22A	AGTCAGGACGCATTGGCTTGCTACCGCTACCGGATCCCGCTCATTTCTAGGCTACCGCTACCAAGAAGAAAGGAAGACATGT
Y22B	AGTCAGGACGCATTGGCTTGTTATCGTTATCGGATCCCGCTCATTTCTAGGTTATCGTTATCAAGAAGAAAGGAAGACATGT
Y23A	AGTCAGGACGCATTGGCTTGGACCCGGACCCGGATCCCGCTCATTTCTAGGGACCCGGACCCAAGAAGAAAGGAAGACATGT
Y23B	AGTCAGGACGCATTGGCTTGTACTCGTACTCGGATCCCGCTCATTTCTAGGTACTCGTACTCAAGAAGAAAGGAAGACATGT
Y24A	AGTCAGGACGCATTGGCTTGTACGGATGTACGGATGGATCCCGCTCATTTCTAGGTACGGATGTACGGATAAGAAGAAAGGAAGACATGT
Y24B	AGTCAGGACGCATTGGCTTGAACGGTTGAACGGTTGGATCCCGCTCATTTCTAGGAACGGTTGAACGGTTAAGAAGAAAGGAAGACATGT
Y25A	AGTCAGGACGCATTGGCTTGTGCACGTGCACGGATCCCGCTCATTTCTAGGTGCACGTGCACAAGAAGAAAGGAAGACATGT
Y25B	AGTCAGGACGCATTGGCTTGTTTACGGTTTACGGGATCCCGCTCATTTCTAGGTTTACGGTTTACGAAGAAGAAAGGAAGACATGT
Y26A	AGTCAGGACGCATTGGCTTTACGTATACGTAGGATCCCGCTCATTTCTAGTACGTATACGTAAAGAAGAAAGGAAGACATGT
Y26B	AGTCAGGACGCATTGGCTTTACCTATACCTAGGATCCCGCTCATTTCTAGTACCTATACCTAAAGAAGAAAGGAAGACATGT
Y27A	AGTCAGGACGCATTGGCTTTAGCCATAGCCAGGATCCCGCTCATTTCTAGTAGCCATAGCCAAAGAAGAAAGGAAGACATGT
Y27B	AGTCAGGACGCATTGGCTTTTGCTATTGCTAGGATCCCGCTCATTTCTAGTTGCTATTGCTAAAGAAGAAAGGAAGACATGT
Y28A	AGTCAGGACGCATTGGCTTTGGCGCTGGCGCGGATCCCGCTCATTTCTAGTGGCGCTGGCGCAAGAAGAAAGGAAGACATGT
Y28B	AGTCAGGACGCATTGGCTTTTGCTCTTGCTCGGATCCCGCTCATTTCTAGTTGCTCTTGCTCAAGAAGAAAGGAAGACATGT
Y29A	AGTCAGGACGCATTGGCTTTGTATGGTGTATGGGGATCCCGCTCATTTCTAGTGTATGGTGTATGGAAGAAGAAAGGAAGACATGT
Y29B	AGTCAGGACGCATTGGCTTTCTGTCGTCTGTCGGGATCCCGCTCATTTCTAGTCTGTCGTCTGTCGAAGAAGAAAGGAAGACATGT
Y30A	AGTCAGGACGCATTGGCTTTGTGGCGTGTGGCGGGATCCCGCTCATTTCTAGTGTGGCGTGTGGCGAAGAAGAAAGGAAGACATGT

Y30B	AGTCAGGACGCATTGGCTTTATGGTGTATGGTGGGATCCCGCTCATTTCTAGTATGGTGTATGGTGAAGAAGAAAGGAAGACATGT
Y31A	AGTCAGGACGCATTGGCTTTGTTCTTGTTCTGGATCCCGCTCATTTCTAGTGTTCTTGTTCTAAGAAGAAAGGAAGACATGT
Y31B	AGTCAGGACGCATTGGCTTAGTTCAAGTTCAGGATCCCGCTCATTTCTAGAGTTCAAGTTCAAAGAAGAAAGGAAGACATGT
Y32A	AGTCAGGACGCATTGGCTTTTCAAGTTCAAGGGATCCCGCTCATTTCTAGTTCAAGTTCAAGAAGAAGAAAGGAAGACATGT
Y32B	AGTCAGGACGCATTGGCTTTTCAGGTTCAGGGGATCCCGCTCATTTCTAGTTCAGGTTCAGGAAGAAGAAAGGAAGACATGT
Y33A	AGTCAGGACGCATTGGCTTTTCCAGAATTCCAGAAGGATCCCGCTCATTTCTAGTTCCAGAATTCCAGAAAAGAAGAAAGGAAGACATGT
Y33B	AGTCAGGACGCATTGGCTTACGAGTATAACGAGTATAGGATCCCGCTCATTTCTAGACGAGTATAAAGAAGAAAGGAAGACATGT
Y34A	AGTCAGGACGCATTGGCTTTTCTTTCTTCTTTCGGATCCCGCTCATTTCTAGTTCTTTCTTCTTTCAAGAAGAAAGGAAGACATGT
Y34B	AGTCAGGACGCATTGGCTTTGCATGCTGCATGCGGATCCCGCTCATTTCTAGTGCATGCTGCATGCAAGAAGAAAGGAAGACATGT
Y35A	AGTCAGGACGCATTGGCTTTGCCACTTGCCACTGGATCCCGCTCATTTCTAGTGCCACTTGCCACTAAGAAGAAAGGAAGACATGT
Y35B	AGTCAGGACGCATTGGCTTTAGCCGCTTAGCCGCTGGATCCCGCTCATTTCTAGTAGCCGCTTAGCCGCTAAGAAGAAAGGAAGACATGT
Y36A	AGTCAGGACGCATTGGCTTTTGGAGTTGGAGGGATCCCGCTCATTTCTAGTTGGAGTTGGAGAAGAAGAAAGGAAGACATGT
Y36B	AGTCAGGACGCATTGGCTTTCGGTGTCGGTCGGTGTCGGGGATCCCGCTCATTTCTAGTCGGTGTCGGTCGGTGTCGGAAGAAGAAAGGAAGACATGT
Y37A	AGTCAGGACGCATTGGCTTTTTCAGATTTCAGAGGATCCCGCTCATTTCTAGTTTCAGATTTCAGAAAGAAGAAAGGAAGACATGT
Y37B	AGTCAGGACGCATTGGCTTTATCACATATCACAGGATCCCGCTCATTTCTAGTATCACATATCACAAAGAAGAAAGGAAGACATGT
Y38A	AGTCAGGACGCATTGGCTTATACCTATACCTGGATCCCGCTCATTTCTAGATACCTATACCTAAGAAGAAAGGAAGACATGT
Y38B	AGTCAGGACGCATTGGCTTAGACTTAGACTTGGATCCCGCTCATTTCTAGAGACTTAGACTTAAGAAGAAAGGAAGACATGT
Y39A	AGTCAGGACGCATTGGCTTGATTGAGATTGAGGATCCCGCTCATTTCTAGGATTGAGATTGAAAGAAGAAAGGAAGACATGT
Y39B	AGTCAGGACGCATTGGCTTGCTTTAGCTTTAGGATCCCGCTCATTTCTAGGCTTTAGCTTTAAAGAAGAAAGGAAGACATGT
Y40A	AGTCAGGACGCATTGGCTTTTGCAATTGCAAGGATCCCGCTCATTTCTAGTTGCAATTGCAAAAGAAGAAAGGAAGACATGT
Y40B	AGTCAGGACGCATTGGCTTTAGCGATAGCGAGGATCCCGCTCATTTCTAGTAGCGATAGCGAAAGAAGAAAGGAAGACATGT

Supplemental Table 4: Table of Probes used in protein microarrays	
P1A	TCGACTCTGGATCCTGGTACGTATGGTACGTATGGTACGTATGGTACGTACATAGTCTACTA
P1B	TCGACTCTGGATCCTAGTGCATATAGTGCATATAGTGCATATAGTGCATACATAGTCTACTA
P2A	TCGACTCTGGATCCTTTCTAGATTTCTAGATTTCTAGATTTCTAGATTCATAGTCTACTA
P2B	TCGACTCTGGATCCTATCTAGATATCTAGATATCTAGATATCTAGATACATAGTCTACTA
P3A	TCGACTCTGGATCCCTACCTCCGAGCTACCTCCGAGCTACCTCCGAGCATAGTCTACTA
P3B	TCGACTCTGGATCCTAGCGATAGCGATAGCGATAGCGATAGCGATAGCATAGTCTACTA
P4A	TCGACTCTGGATCCACCAAGCAAAAACCAAGCAAAAACCAAGCAAAACATAGTCTACTA
P4B	TCGACTCTGGATCCACGAACCATAAACGAACCATAAACGAACCATAACATAGTCTACTA
P5A	TCGACTCTGGATCCAACAACAACAACAACAACAACAACAACAACAACCCATAGTCTACTA
P5B	TCGACTCTGGATCCAATGACAATGACAATGACAATGACAATGACAACCCATAGTCTACTA
P6A	TCGACTCTGGATCCAACTTTTCAACTTTTCAACTTTTCAACTTTTCCCATAGTCTACTA
P6B	TCGACTCTGGATCCTTCTTTTGTTCTTTTGTTCTTTTGTTCTTTTCCCATAGTCTACTA
P7A	TCGACTCTGGATCCAAGCCACAAAGCCACAAAGCCACACCACACATAGTCTACTA
P7B	TCGACTCTGGATCCAACTCACAAACTCACAAACTCACACCACACATAGTCTACTA
P8A	TCGACTCTGGATCCACATACACATACACATACACATACACATACACATCATAGTCTACTA
P8B	TCGACTCTGGATCCAGTTGCAGTTGCAGTTGCAGTTGCAGTTGAGTTGCATAGTCTACTA
P9A	TCGACTCTGGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCATAGTCTACTA
P9B	TCGACTCTGGATCCACACTGACACTGACACTGACACTGACACTGACCATAGTCTACTA
P10A	TCGACTCTGGATCCAGTGAAAGTGAAAGTGAAAGTGAAAGTGAAAGTGCATAGTCTACTA
P10B	TCGACTCTGGATCCAGTCAAAGTCAAAGTCAAAGTCAAAGTCAAAGTCCATAGTCTACTA
P11A	TCGACTCTGGATCCATATGTATATGTATATGTATATGTATATGTATATCATAGTCTACTA
P11B	TCGACTCTGGATCCAGATTGAGATTGAGATTGAGATTGAGATTGAGATCATAGTCTACTA
P12A	TCGACTCTGGATCCATGCGATGATGCGATGATGCGATGATGCGATGATCATAGTCTACTA
P12B	TCGACTCTGGATCCACGCGAAGACGCGAAGACGCGAAGACGCGAAGACCATAGTCTACTA
P13A	TCGACTCTGGATCCCAAACAAACAAACAAACAAACAAACAAACAAACCATAGTCTACTA
P13B	TCGACTCTGGATCCCAGACATACAGACATACAGACATACAGACATACCATAGTCTACTA

P14A	TCGACTCTGGATCCCAAGGGCAAGGGCAAGGGCAAGGGCAAGGGCAAGCATAGTCTACTA
P14B	TCGACTCTGGATCCCAAGTGCAAGTGCAAGTGCAAGTGCAAGTGCAAGCATAGTCTACTA
P15A	TCGACTCTGGATCCCACCACCACCACCACCACCACCACCACCACCACCCATAGTCTACTA
P15B	TCGACTCTGGATCCCACTACCACTACCACTACCACTACCACTACCACCCATAGTCTACTA
P16A	TCGACTCTGGATCCAATAACAATAACAATAACAATAACAATAACAACCCATAGTCTACTA
P16B	TCGACTCTGGATCCCGACACCGACACCGACACCGACACCGACACCGACCATAGTCTACTA
P17A	TCGACTCTGGATCCCGATACCTACACCTACACCTACACCTACACCTACCATAGTCTACTA
P17B	TCGACTCTGGATCCCTCACGCTCACGCTCACGCTCACGCTCACGCTCACATAGTCTACTA
P18A	TCGACTCTGGATCCCTAAACGACTAAACGACTAAACGACTAAACGCATAGTCTACTA
P18B	TCGACTCTGGATCCCGATACTACGATACTACGATACTACGATACTCATAGTCTACTA
P19A	TCGACTCTGGATCCCTGAAAACTGAAAACTGAAAACTGAAAACATAGTCTACTA
P19B	TCGACTCTGGATCCCAGAATACAGAATACAGAATACAGAATACATAGTCTACTA
P20A	TCGACTCTGGATCCGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCACATAGTCTACTA
P20B	TCGACTCTGGATCCGCTAACGCTAACGCTAACGCTAACGCTAACGCTACATAGTCTACTA
P21A	TCGACTCTGGATCCGCGATGCGCGATGCGCGATGCGCGATGCGCGATCATAGTCTACTA
P21B	TCGACTCTGGATCCGTGATTCGTGATTCGTGATTCGTGATTCGTGATCATAGTCTACTA
P22A	TCGACTCTGGATCCGCTACCGCTACCGCTACCGCTACCGCTACCGCTACATAGTCTACTA
P22B	TCGACTCTGGATCCGTTATCGTTATCGTTATCGTTATCGTTATCGTTACATAGTCTACTA
P23A	TCGACTCTGGATCCGGACCCGGACCCGGACCCGGACCCGGACCCGGACCATAGTCTACTA
P23B	TCGACTCTGGATCCGTACTCGTACTCGTACTCGTACTCGTACTCGTACCATAGTCTACTA
P24A	TCGACTCTGGATCCGTACGGATGTACGGATGTACGGATGTACGGATCATAGTCTACTA
P24B	TCGACTCTGGATCCGAACGGTTGAACGGTTGAACGGTTGAACGGTTCATAGTCTACTA
P25A	TCGACTCTGGATCCGTGCACGTGCACGTGCACGTGCACGTGCACGTGCCATAGTCTACTA
P25B	TCGACTCTGGATCCGTTTACGTTTACGTTTACGTTTACGTTTACGTTTCATAGTCTACTA
P26A	TCGACTCTGGATCCTACGTATACGTATACGTATACGTATACGTATACGCATAGTCTACTA
P26B	TCGACTCTGGATCCTTACCTATACCTATACCTATACCTATACCTATACCATAGTCTACTA
P27A	TCGACTCTGGATCCTAGCCATAGCCATAGCCATAGCCATAGCCATAGCCATAGTCTACTA
P27B	TCGACTCTGGATCCTTGCTATTGCTATTGCTATTGCTATTGCTATTGCCATAGTCTACTA

P28A	TCGACTCTGGATCCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCCATAGTCTACTA
P28B	TCGACTCTGGATCCTTGCTCTTGCTCTTGCTCTTGCTCTTGCTCTTGCCATAGTCTACTA
P29A	TCGACTCTGGATCCTGTATGGTGTATGGTGTATGGTGTATGGTGTATGCATAGTCTACTA
P29B	TCGACTCTGGATCCTCTGTCGTCTGTCGTCTGTCGTCTGTCGTCTGTCCATAGTCTACTA
P30A	TCGACTCTGGATCCTGTGGCGTGTGGCGTGTGGCGTGTGGCGTGTGGCCATAGTCTACTA
P30B	TCGACTCTGGATCCTATGGTGTATGGTGTATGGTGTATGGTGTATGGTCATAGTCTACTA
P31A	TCGACTCTGGATCCTGTTCTTGTTCTTGTTCTTGTTCTTGTTCTTGTTCATAGTCTACTA
P31B	TCGACTCTGGATCCAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGCACATAGTCTACTA
P32A	TCGACTCTGGATCCTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCACATAGTCTACTA
P32B	TCGACTCTGGATCCTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCACATAGTCTACTA
P33A	TCGACTCTGGATCCTTCCAGAATTCCAGAATTCCAGAATTCCAGAATTCATAGTCTACTA
P33B	TCGACTCTGGATCCTTCCTACGAGTATACGAGTATACGAGTATACGAGCATAGTCTACTA
P34A	TCGACTCTGGATCCTTCTTTCTTCTTTCTTCTTTCTTCTTTCTTCTTTCATAGTCTACTA
P34B	TCGACTCTGGATCCTGCATGCTGCATGCTGCATGCTGCATGCTGCATGCATAGTCTACTA
P35A	TCGACTCTGGATCCTTGCCACTTTGCCACTTTGCCACTTTGCCACTTTCATAGTCTACTA
P35B	TCGACTCTGGATCCTAGCCGCTTAGCCGCTTAGCCGCTTAGCCGCTTACATAGTCTACTA
P36A	TCGACTCTGGATCCTTGGAGTTGGAGTTGGAGTTGGAGTTGGAGTTGGCATAGTCTACTA
P36B	TCGACTCTGGATCCTCGGTGTCGGTGTCGGTGTCGGTGTCGGTGTCGGCATAGTCTACTA
P37A	TCGACTCTGGATCCTTTCAGATTTCAGATTTCAGATTTCAGATTTCAGCATAGTCTACTA
P37B	TCGACTCTGGATCCTATCACATATCACATATCACATATCACATATCACCATAGTCTACTA
P38A	TCGACTCTGGATCCATACCTATACCTATACCTATACCTATACCTATACATAGTCTACTA
P38B	TCGACTCTGGATCCAGACTTAGACTTAGACTTAGACTTAGACTTAGACATAGTCTACTA
P39A	TCGACTCTGGATCCGATTGAGATTGAGATTGAGATTGAGATTGAGATTCATAGTCTACTA
P39B	TCGACTCTGGATCCGCTTTAGCTTTAGCTTTAGCTTTAGCTTTAGCTTCATAGTCTACTA
P40A	TCGACTCTGGATCCTTGCAATTGCAATTGCAATTGCAATTGCAATTGCCATAGTCTACTA
P40B	TCGACTCTGGATCCTAGCGATAGCGATAGCGATAGCGATAGCGATAGCCATAGTCTACTA
P41A	GNNNGCGATAGNNNGCGATAGNNNGCGATAGNNNCATAGTCTACTA
P41B	GNNNGTGCTAGNNNGTGCTAGNNNGTGCTAGNNNCATAGTCTACTA

P42A	GNNNGTGGCGNNNGTGGCGNNNGTGGCGNNNCATAGTCTACTA
P42B	GNNNGTAGTGNNNGTAGTGNNNGTAGTGNNNCATAGTCTACTA
P43	TCGACTCTGGATCCCCGAGCTACCTCCGCATAGTCTACTA
P44	TCGACTCTGGATCCACGAGCTACCTCCGCATAGTCTACTA
P45	TCGACTCTGGATCCCAGAGCTACCTCCGCATAGTCTACTA
P46	TCGACTCTGGATCCCCCAGCTACCTCCGCATAGTCTACTA
P47	TCGACTCTGGATCCCCGAGCTACCTACGCATAGTCTACTA
P48	TCGACTCTGGATCCCCGAGCTACCTCAGCATAGTCTACTA
P49	TCGACTCTGGATCCCCGAGCTACCTCCCCATAGTCTACTA
P50	TCGACTCTGGATCCCAGAGCTACCTCAGCATAGTCTACTA
P51	TCGACTCTGGATCCCCGAGCTCCTCCGCATAGTCTACTA
P52	TCGACTCTGGATCCCCGAGCTAGCCTCCGCATAGTCTACTA

Supplemental Table 1: Probes sequences used in protein microarrays and EMSA. P1A through P40B are the pairs of probes used to probe the protein microarrays. P41 and P42 were designed to take advantage of the binding logos predicted for Stp3 and Yml081W respectively, and used in EMSA. P43-P52 are the panel of probes tested for Yjl103c binding.

Supplemental Table 5: Candidates from protein microarrays

Probe	ORF	Protein
P1A	YDL170W	UGA3
P1A	YLR228C	ECM22
P1A	YML076C	WAR1
P1A	YNL167C	SKO1
P1A	YOR028C	CIN5
P1A	YOR344C	TYE7
P1A	YPR008W	HAA1
P3A	YAL051W	OAF1
P3A	YGL035C	MIG1
P3A	YJL103C	YJL103C
P3A	YKL038W	RGT1
P3A	YOR162C	YRR1
P3A	YOR380W	RDR1
P3B	YBR112C	CYC8
P3B	YDL048C	STP4
P3B	YGL254W	FZF1
P3B	YHL009C	YAP3
P3B	YHR056C	RSC30
P3B	YHR206W	SKN7
P3B	YIL036W	CST6
P3B	YJL110C	GZF3
P3B	YKL112W	ABF1
P3B	YLR176C	RFX1
P3B	YLR375W	STP3
P3B	YNL167C	SKO1
P4A	YDL048C	STP4
P4A	YDR034C	LYS14
P4A	YKL043W	PHD1
P4A	YMR280C	CAT8
P7A	YBL021C	HAP3
P7A	YDR253C	MET32
P7A	YHR206W	SKN7
P7A	YIL036W	CST6
P7A	YKL043W	PHD1
P7A	YNL027W	CRZ1
P7A	YOR032C	HMS1
P7A	YPL038W	MET31
P7B	YBL021C	HAP3
P7B	YCR047C	BUD23
P22B	YOR380W	RDR1
P24A	YAL051W	OAF1
P24A	YNL167C	SKO1
P24A	YNL216W	RAP1
P24A	YOR380W	RDR1
P25A	YML081W	YML081W
P25A	YOR344C	TYE7
P26A	YAL034W-A	MTW1
P26A	YBL021C	HAP3
P26A	YDR213W	UPC2
P26A	YKL043W	PHD1
P27A	YDR253C	MET32
P27A	YHR206W	SKN7
P27A	YLR375W	STP3
P28A	YDL170W	UGA3
P28A	YGL209W	MIG2
P28A	YHR206W	SKN7
P28A	YLR098C	CHA4
P28A	YLR375W	STP3
P28A	YNL167C	SKO1
P28A	YNL216W	RAP1
P28B	YKR099W	BAS1
P28B	YMR037C	MSN2
P28B	YMR039C	SUB1
P29A	YKR099W	BAS1
P29A	YMR039C	SUB1
P29A	YMR168C	CEP3
P29A	YNL216W	RAP1
P29B	YFR034C	PHO4
P30A	YDL125C	HNT1
P30A	YDR253C	MET32
P30A	YFR034W	PHO4
P30A	YLR131C	ACE2
P30A	YLR375W	STP3
P30A	YML081W	YML081W
P30A	YMR072W	ABF2
P30A	YMR168C	CEP3
P30A	YNL216W	RAP1
P30A	YOR032C	HMS1

P7B	YOR032C	HMS1
P8A	YCR096C	HMRA2
P8A	YDL125C	HNT1
P8A	YDR253C	MET32
P8A	YFR034W	PHO4
P8A	YKL043W	PHD1
P8A	YMR168C	CEP3
P8A	YPL038W	MET31
P8A	YNL167C	SKO1
P8A	YER111C	SWI4
P8B	YKL043W	PHD1
P8B	YPR008W	HAA1
P9B	YDR253C	MET32
P9B	YKL043W	PHD1
P9B	YKR099W	BAS1
P9B	YMR039C	SUB1
P9B	YMR182C	RGM1
P10A	YBL021C	HAP3
P10A	YKL043W	PHD1
P10A	YKR099W	BAS1
P11A	YCR096C	HMRA2
P11A	YKL043W	PHD1
P11A	YPR008W	HAA1
P16A	YCR096C	HMRA2
P16A	YKL020C	SPT23
P16B	YKL043W	PHD1
P17A	YHL009C	YAP3
P17A	YNL167C	SKO1
P18A	YDR213W	UPC2
P18A	YER111C	SWI4
P18A	YHL009C	YAP3
P18A	YLR228C	ECM22
P18B	YMR168C	CEP3
P18B	YDL125C	HNT1
P19A	YPR008W	HAA1
P19A	YER111C	SWI4
P19A	YHL009C	YAP3
P19A	YKL043W	PHD1
P19A	YKR048C	NAP1
P19A	YML076C	WAR1
P19A	YMR037C	MSN2
P19A	YMR039C	SUB1

P30A	YOR344C	TYE7
P30A	YPL038W	MET31
P30A	YPL133C	RDS2
P30B	YGL254W	FZF1
P30B	YIL036W	CST6
P30B	YJL110C	GZF3
P30B	YKL043W	PHD1
P30B	YKR099W	BAS1
P30B	YLR228C	ECM22
P30B	yml076C	WAR1
P30B	YML081W	YML081W
P30B	YNL167C	SKO1
P30B	YNL216W	RAP1
P30B	YNL227C	YNL227C
P30B	YOL028C	YAP7
P30B	YOR032C	HMS1
P30B	YOR113W	AZF1
P30B	YOR344C	TYE7
P30B	YOR380W	RDR1
P33A	YMR072W	ABF2
P33A	YCR096C	HMRA2
P33A	YDL170W	UGA3
P33A	YDR034C	LYS14
P33A	YGL035C	MIG1
P33A	YGL073W	HSF1
P33A	YGR249W	MGA1
P33A	YHR206W	SKN7
P33A	YJR089W	BIR1
P33A	YKR048C	NAP1
P33A	YKR099W	BAS1
P33A	YLR228C	ECM22
P33A	YML010W	SPT5
P33A	YML081W	YML081W
P33A	YMR037C	MSN2
P33A	YMR039C	SUB1
P33A	YMR280C	CAT8
P33A	YOL028C	YAP7
P33A	YOR113W	AZF1
P33A	YOR380W	RDR1
P33A	YPR065W	ROX1
P33B	YCR096C	HMRA2
P33B	YAL034W-A	MTW1

P19A	YMR043W	MCM1
P19A	YOR162C	YRR1
P19A	YOR380W	RDR1
P19B	YAL034W-A	MTW1
P20A	YBL021C	HAP3
P20A	YCR096C	HMRA2
P20A	YDR009W	GAL3
P20A	YHR206W	SKN7
P20A	YNL027W	CRZ1
P21A	YDL170W	UGA3
P21A	YDR303C	RSC3
P21A	YKL043W	PHD1
P21A	YLR098C	CHA4
P21A	YML076C	WAR1
P21A	YOR380W	RDR1
P21B	YMR168C	CEP3
P21B	YDL125C	HNT1
P21B	YOR337W	TEA1
P22A	YDL125C	HNT1
P22A	YJL110C	GZF3
P22A	YKL112W	ABF1
P22A	YLR228C	ECM22
P22B	YKL043W	PHD1
P22B	YML076C	WAR1

P33B	YER111C	SWI4
P35A	YDR253C	MET32
P35A	YKR048C	NAP1
P35A	YML027W	YOX1
P35A	YMR039C	SUB1
P35B	YOR380W	RDR1
P36A	YGL035C	MIG1
P37A	YHR206W	SKN7
P38A	YCR084C	TUP1
P38A	YMR039C	SUB1
P38A	YMR280C	CAT8
P38A	YNL167C	SKO1
P38A	YPR196W	YPR196W
P39A	YJL110C	GZF3
P39B	YOR380W	RDR1
P40A	YDR323C	PEP7
P40A	YLR131C	ACE2
P40A	YLR204W	QRI5
P40A	YML076C	WAR1
P40A	YML081W	YML081W
P40A	YMR280C	CAT8
P40A	YNL227C	YNL227C
P40A	YOL028C	YAP7
P40B	YHR206W	SKN7

Supplemental Table 6: Proteins identified by Maldi-tof for both Sef1 and Yjl103c	
Adh1	Alcohol dehydrogenase, fermentative isozyme active as homo- or heterotetramers; required for the reduction of acetaldehyde to ethanol, the last step in the glycolytic pathway
Bmh1	14-3-3 protein, major isoform; controls proteome at post-transcriptional level, binds proteins and DNA, involved in regulation of many processes including exocytosis, vesicle transport, Ras/MAPK signaling, and rapamycin-sensitive signaling
Cit1	Citrate synthase, catalyzes the condensation of acetyl coenzyme A and oxaloacetate to form citrate; the rate-limiting enzyme of the TCA cycle;
Eno1	Enolase I, a phosphopyruvate hydratase that catalyzes the conversion of 2-phosphoglycerate to phosphoenolpyruvate during glycolysis and the reverse reaction during gluconeogenesis; expression is repressed in response to glucose
Eno2	Enolase II, a phosphopyruvate hydratase that catalyzes the conversion of 2-phosphoglycerate to phosphoenolpyruvate during glycolysis and the reverse reaction during gluconeogenesis; expression is induced in response to glucose
Fba1	Fructose 1,6-bisphosphate aldolase, required for glycolysis and gluconeogenesis; catalyzes conversion of fructose 1,6 bisphosphate to glyceraldehyde-3-P and dihydroxyacetone-P;

Grx1	Hydroperoxide and superoxide-radical responsive heat-stable glutathione-dependent disulfide oxidoreductase with active site cysteine pair; protects cells from oxidative damage
Pgk1	3-phosphoglycerate kinase, catalyzes transfer of high-energy phosphoryl groups from the acyl phosphate of 1,3-bisphosphoglycerate to ADP to produce ATP; key enzyme in glycolysis and gluconeogenesis
Tdh3	Glyceraldehyde-3-phosphate dehydrogenase, isozyme 3, involved in glycolysis and gluconeogenesis.
Ura3	Orotidine-5'-phosphate (OMP) decarboxylase, catalyzes the sixth enzymatic step in the de novo biosynthesis of pyrimidines, converting OMP into uridine monophosphate (UMP).

Supplemental Table 7: Yeast strains used. Strains not created for this study are referenced			
Strain	Genotype	function	Reference
FM391	*MATa his3Δ leu2Δ ura3Δ met15Δ* (BY4741)	Wild-type Mata strain	(21)
FM392	*MATα his3Δ leu2Δ ura3Δ lys2Δ* (BY4742)	Wild-type *MAT*α strain	(21)
FM393	*MATa/MATα his3Δ/his3Δ leu2Δ/leu2Δ ura3Δ/ura3Δ met15Δ/MET LYS/lys2Δ* (BY4743)	Wild-type *MATa/MAT*α strain	(21)
PJ69-4a	(*MATa trp1-901 leu2-3,112 ura3-52 his3-200 gal4 gal80 LYS2::GAL1-HIS3 GAL2-ADE2 met2::GAL7-lacZ*)	Yeast 1-hybrid AD fusion	(95)
PJ69-4α	(*MATα trp1-901 leu2-3,112 ura3-52 his3-200 gal4 gal80 LYS2::GAL1-HIS3 GAL2-ADE2 met2::GAL7-lacZ*)	Yest 1-hybrid *HIS3* reporter strain	(208)
YM7374	*MATa his3Δ leu2Δ ura3Δ met15Δ*	Overexpressed *SEF1-GST*	
Deletions	*MATa his3Δ leu2Δ ura3Δ met15Δ* (BY4741)	*Yjl103cΔ Sef1Δ*	(63)
Y258	*MATa ga14-542Δura3-his3-200Δ ade2Δ lys2-801Δ trp1-901Δ tyr1-501*	GST-fusion proteins	(95)

Plasmids containing TF AD fusions were given as a gift from Stan Fields, University of Washington. Strains expressing GST fusion proteins were given as a gift from Michael Snyder, Yale University.

References

References:

1. **Akache, B., S. MacPherson, M. A. Sylvain, and B. Turcotte.** 2004. Complex interplay among regulators of drug resistance genes in Saccharomyces cerevisiae. J. Biol. Chem. **279:**27855-60.
2. **Alam J., and J. L. Cook.** 1990. Reporter genes: application to the study of mammalian gene transcription. Anal. Biochem. **188:**245-54.
3. **Alexander, M. K., B. D. Bourns, and V.A. Zakian.** 2001. One-hybrid systems for detecting protein–DNA interactions. Methods Molecular Biology **177:**241-259.
4. **Angus-Hill ML, S. A., D. Roberts, H. Erdjument-Bromage, P. Tempst,and B. R. Cairns.** 2001. A Rsc3/Rsc30 zinc cluster dimer reveals novel roles for the chromatin remodeler RSC in gene expression and cell cycle control. Mol. Cell **7:**741-5.
5. **Aravind, L., V. Anantharaman, S. Balaji, M. M. Babu, and L. M. Iyer.** 2005. The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev **29:**231-62.
6. **Avram D., M. Leid, and A. T. Bakalinsky.** 1999. Fzf1p of Saccharomyces cerevisiae is a positive regulator of SSU1 transcription and its first zinc finger region is required for DNA binding. Yeast. **15:**473-80.
7. **Axelrod J. D., J. Majors, and M. C. Brandriss.** 1991. Proline-independent binding of PUT3 transcriptional activator protein detected by footprinting in vivo. Mol. Cell Biol. **11:**564-7.

8. **Badis G., E. T. Chan, H. van Bakel, L. Pena-Castillo, D. Tillo, K. Tsui, C. D. Carlson, A. J. Gossett, M. J. Hasinoff, C. L. Warren, M. Gebbia, S. Talukder, A. Yang, S. Mnaimneh, D. Terterov, D. Coburn, A. Li Yeo , Z. X. Yeo, N. D. Clarke, J. D. Lieb, A. Z. Ansari, C. Nislow, and T. R. Hughes.** 2008. A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol. Cell **32:**878-87.

9. **Bailey T. L., and C. Elkan.** 1995. The value of prior knowledge in discovering motifs with MEME. Pro. Int. Conf. Intell. Syst Mol. Biol. **3:**21-9.

10. **Baleja, J. D., R. Marmorstein, S. C. Harrison, and G. Wagner.** 1992. Solution structure of the DNA-binding domain of Cd2-GAL4 from S. cerevisiae. Nature **356:**450-3.

11. **Bejerano G., M. Pheasant, I. Makunin, S. Stephen, W. J. KenT, J. S. Mattick, D. Haussler.** 2004. Ultraconserved elements in the human genome. Science **304:**1321-5.

12. B**enos P.V., M. L. Bulyk, and G. D. Stormo.** 2002. Additivity in protein-DNA interactions: how good an approximation is it? Nucleic Acids Res. **30:**4442-51.

13. **Benos, P. V., A. S. Lapedes, and G. D. Stormo.** 2002. Is there a code for protein-DNA recognition? Probab(ilistical)ly. Bioessays **24:**466-75.

14. **Benos, P. V., A. S. Lapedes, and G. D. Stormo.** 2002. Probabilistic code for DNA recognition by proteins of the EGR family. J. Mol. Biol. **323:**701-27.

15. **Blaiseau P. L., A. D. Isnard, Y. Surdin-Kerjan, D. Thomas.** 1997. Met31p and Met32p, two related zinc finger proteins, are involved in transcriptional regulation of yeast sulfur amino acid metabolism. Mol. Cell Biol. **17:**3640-8.

16. **Borneman A. R., J. A. Leigh-Bell, H. Yu, P. Bertone, M. Gerstein, and M. Snyder.** 2006. Target hub proteins serve as master regulators of development in yeast. Genes Dev. **20:**435-48.

17. **Bourges I., M. H. Mucchielli, C. J. Herbert, B. Guiard, G. Dujardin, and B. Meunier.** 2009. Multiple defects in the respiratory chain lead to the repression of genes encoding components of the respiratory chain and TCA cycle enzymes. J. Mol. Biol. **387:**1081-91.

18. **Bouvet P.** 2008. Identification of Nucleic Acid High-Affinity Binding Sequences of Proteins by SELEX. Methods in Molecular Biology **543:**139-150.

19. **Bowen B., J. Steinberg, U. K. Laemmli, and H. Weintraub.** 1980. The detection of DNA-binding proteins by protein blotting. Nucleic Acids Res. **8:**1-20

20. **Boyer L. A. , T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, and R. A. Young.** 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell **122:**947-56.

21. **Brachmann C. B., A. Davis, G. J. Cost, E. Caputo, J. Li , P. Hieter, and J. D. Boeke.** 1998. Designer deletion strains derived from Saccharomyces

cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast **14:**115-23.

22. **Brent, R., and M. Ptashne.** 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell **43:**729-36.

23. **Bricmont P. A., J. R. Daugherty, and Cooper TG.** 1991. The DAL81 gene product is required for induced expression of two differently regulated nitrogen catabolic genes in Saccharomyces cerevisiae. Mol. Cell Biol. **11:**1161-6.

24. **Brown, C. E., T. Lechner, L. Howe, and J. L. Workman.** 2000. The many HATs of transcription coactivators. Trends Biochem Sci **25:**15-9.

25. **Bulyk M. L., P. L. Johnson, and G. M. Church.** 2002. Nucleotides of transcription factor binding sites exert interdependent effects on the binding affinities of transcription factors. Nucleic Acids Res. **30:**1255-61.

26. **Bun-Ya M., M. Nishimura, S. Harashima, and Y. Oshima.** 1991. The PHO84 gene of Saccharomyces cerevisiae encodes an inorganic phosphate transporter. Mol. Cell Biol. **11:**3229-38.

27. **Buschlen S., J. M. Amillet, B. Guiard, A. Fournier, C. Marcireau, and M. Bolotin-Fukuhara.** 2003. The S. Cerevisiae HAP Complex, a Key Regulator of Mitochondrial Function, Coordinates Nuclear and Mitochondrial Gene Expression. Comp. Funct. Genomics. **4:**37-46.

28. **Chen G., L. DenBoer, and J. Shin.** 2008. Design of a single plasmid-based modified yeast one-hybrid system for investigation of in vivo protein-protein and protein-DNA interactions. Biotechniques **45:**295-304.

29. **Chen J., M. Ding, and D. S. Pederson.** 1994. Binding of TFIID to the CYC1 TATA boxes in yeast occurs independently of upstream activating sequences. Proc. Natl. Acad. Sci. U S A. **91:**11909-13.

30. **Christopoulos, T. K., E. P. Diamandis, and G. Wilson.** 1991. Quantification of nucleic acids on nitrocellulose membranes with time-resolved fluorometry. Nucleic Acids Res. **19:**6015-6019.

31. **Cliften P, P. Sudarsanam, A. Desikan, L. Fulton, B. Fulton, J. Majors, R. Waterston, B. A. Cohen, and M. Johnston.** 2003. Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science **301:**71-6.

32. **Cliften, P. F., L. W. Hillier, L. Fulton, T. Graves, T. Miner, W. R. Gish, R. H. Waterston, and M. Johnston.** 2001. Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis. Genome Res **11:**1175-86.

33. **Costanzo M., A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear, C. S. Sevier, H. Ding, J. L. Koh, K. Toufighi, S. Mostafavi, J. Prinz,R. P. St Onge, B. VanderSluis, T. Makhnevych, F. J. Vizeacoumar, S. Alizadeh, S. Bahr, R. L. Brost, Y. Chen, M. Cokol, R. Deshpande, Z. Li, Z. Y. Lin, W. Liang, M. Marback, J. Paw, B. J. San Luis, E. Shuteriqi, A. H. Tong, N. van Dyk, I. M. Wallace, J. A. Whitney, M. T. Weirauch, G. Zhong, H. Zhu, W. A. Houry, M. Brudno, S. Ragibizadeh, B. Papp, C. Pál, F. P. Roth, G. Giaever, C. Nislow, O. G. Troyanskaya, H. Bussey, G. D. Bader, A. C. Gingras, Q. D. Morris, P. M. Kim, C. A. Kaiser, C. L. Myers, B. J.**

Andrews, and C. Boone. 2010. The genetic landscape of a cell. Science **327:**425-31.

34. **Crick, F.** 1970. Central dogma of molecular biology. Nature **227:**561-3.

35. **Crick, F. H.** 1958. On protein synthesis. Symp Soc Exp Biol **12:**138-63.

36. **Crooks G. E. , G. Hon, J. M. Chandonia, and S. E. Brenner.** 2004. WebLogo: a sequence logo generator. Genome Res. **14:**1188-90.

37. **Czerwinski, J. D., S. C. Hovan, and D. P. Mascotti.** 2005. Quantitative nonisotopic nitrocellulose filter binding assays: bacterial manganese superoxide dismutase–DNA interactions. Anal. Biochem. **336:**300-304.

38. **Daniel J. A., J. Yoo, B. T. Bettinger, D. C. Amberg, and D. J. Burke.** 2006. Eliminating gene conversion improves high-throughput genetics in Saccharomyces cerevisiae. Genetics **172:**709-11.

39. **Daran-Lapujade P., M. L. Jansen, J. M. Daran, W. van Gulik W, J. H. de Winde, and J. T. Pronk.** 2004. Role of transcriptional regulation in controlling fluxes in central carbon metabolism of Saccharomyces cerevisiae. A chemostat culture study. J. Biol. Chem. **279:**9125-38.

40. **Davis M. A. , A. J. Small, S. Kourambas, and M. J. Hynes.** 1996. The tamA gene of Aspergillus nidulans contains a putative zinc cluster motif which is not required for gene function. J. Bacteriol. **178:**3406-9.

41. **Deng, Y., T. He, Y. Wu, P. Vanka, G. Yang, Y. Huang, H. Yao, and S. J. Brown.** 2005. Computationally analyzing the possible biological function of YJL103C--an ORF potentially involved in the regulation of energy process in yeast. Int. J. Mol. Med. **15:**123-7.

42. **Deplancke B., D. Dupuy, M. Vidal, and A. J. Walhout.** 2004. A gateway-compatible yeast one-hybrid system. Genome Res. **14:**2093-101.

43. **Deplancke B., A. Mukhopadhyay, W. Ao, A. M. Elewa, C. A. Grove, N. J. Martinez, R. Sequerra, L. Doucette-Stamm, J. S. Reece-Hoyes, I. A. Hope, H. A. Tissenbaum, S. E. Mango, and A. J. Walhout.** 2006. A gene-centered C. elegans protein-DNA interaction network. Cell **125:**1193-205.

44. **Deppmann, C. D., R. S. Alvania, and E. J. Taparowsky.** 2006. Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks. Mol. Biol. Evol. **23:**1480-92.

45. **DeRisi J. L., V. R. Iyer, and P. O. Brown.** 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science **278:**680-6.

46. **Drouin R., N. Bastein, J. Millau, F. Vigneualt, and I. Paradis.** 2008. In Cellulo DNA Analysis (LMPCR Footprinting). Methods in Molecular Biology **453:**293-336.

47. **Dudley A. M., J. Aach, M. A. Steffen, and G. M. Church.** 2002. Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. Proc. Natl. Acad. Sci. U S A. **99:**7554-9.

48. **Elnitski L., V. X. Jin, P. J. Farnham, and S. J. Jones.** 2006. Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. Genome Res. **16:**1455-64.

49. **Fields S., and O. Song.** 1989. A novel genetic system to detect protein-protein interactions. Nature**:**6230.

50. **Finley R. L. Jr, S. Chen, J. Ma, P. Byrne, and R. W. Jr West.** 1990. Opposing regulatory functions of positive and negative elements in UASG control transcription of the yeast GAL genes. Mol. Cell Biol. **10:**5663-70.

51. **Flick J. S., and J. Thorner** 1998. An essential function of a phosphoinositide-specific phospholipase C is relieved by inhibition of a cyclin-dependent protein kinase in the yeast Saccharomyces cerevisiae. Genetics **148:**33-47.

52. **Forsburg S. L. , and L. Guarente.** 1989. Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer. Genes Dev. **3:**1166-78.

53. **Frankel, A. D., J. M. Berg, and C. O. Pabo.** 1987. Metal-dependent folding of a single zinc finger from transcription factor IIIA. Proc. Natl. Acad. Sci. U S A **84:**4841-5.

54. **Fried M., and D. M. Crothers.** 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9:**6505-25.

55. **Galas, D.J. and A. Schmitz.** 1978. DNAse footprinting: a simple method for the detection of protein–DNA binding specificity. Nucleic Acids Res. **5:**3157-70.

56. **Ganter, B., S. Tan, and T. J. Richmond.** 1993. Genomic footprinting of the promoter regions of STE2 and STE3 genes in the yeast Saccharomyces cerevisiae. J. Mol. Biol. **234:**975-87.

57. **Gardner, K. H., T. Pan, S. Narula, E. Rivera, and J. E. Coleman.** 1991. Structure of the binuclear metal-binding site in the GAL4 transcription factor. Biochemistry **30:**11292-302.

58. **Garner M. M., and A. Revzin.** 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic Acids. Res. **9:**3047-60.

59. **Gershenzon N. I. , G. D. Stormo, and I. P. Ioshikhes.** 2005. Computational technique for improvement of the position-weight matrices for the DNA/protein binding sites. Nucleic Acids Res. **33:**2290-301.

60. **Gertz J., and B. A. Cohen.** 2009. Environment-specific combinatorial cis-regulation in synthetic promoters. Mol. Syst. Biol. **5:**244.

61. **Gertz J., L. Riles, P. Turnbaugh, S. W. Ho, and B. A. Cohen.** 2005. Discovery, validation, and genetic dissection of transcription factor binding sites by comparative and functional genomics. Genome Res. **15:**1145-52.

62. **Gertz J, S. E., and B. A. Cohen.** 2009. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature **457:**215-8.

63. **Giaever G., A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Véronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. André, A. P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K. D. Entian, P. Flaherty, F. Foury,D. J. Garfinkel, M. Gerstein, D. Gotte, U. Güldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L.**

Kelly, P. Kötter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C. Y. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen,R. W. Davis, and M. Johnston. 2002. Functional profiling of the Saccharomyces cerevisiae genome. Nature. **418:**387-91.

64. **Gietz R. D., and R. A. Woods.** 2006. Yeast transformation by the LiAc/SS Carrier DNA/PEG method. Methods Mol. Biol. **313**.

65. **Giots F., M. Donaton, and J. M. Thevelein.** 2003. Inorganic phosphate is sensed by specific phosphate carriers and acts in concert with glucose as a nutrient signal for activation of the protein kinase A pathway in the yeast Saccharomyces cerevisiae. Mol. Microbiol. **47:**1163-81.

66. **Glazov E. A., M. Pheasant, E. A. McGraw, G. Bejerano, and J. S. Mattick.** 2005. Ultraconserved elements in insect genomes: a highly conserved intronic sequence implicated in the control of homothorax mRNA splicing. Genome Res. **16:**800-8.

67. **Goffeau A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver.** 1996. Life with 6000 genes. Science **274:**563-7.

68. **Gold, L., B. Polisky, O. Uhlenbeck, and M. Yarus.** 1995. Diversity of oligonucleotide functions. Annu. Rev. Biochem. **64:**763-797.

69. **Gong W., K. Covington, S. P. Dinesh-Kumar, M. Snyder, S. L. Harmer, Y.-X. Zhu and X. W. Deng.** 2008. The Development of Protein Microarrays and Their Applications in DNA–Protein and Protein–Protein Interaction Analyses of Arabidopsis Transcription Factors. Mol. Plant **1:**27-41.

70. **Gong Y., Y. Kakihara, N. Krogan, J. Greenblatt, A. Emili, Z. Zhang, and W. A. Houry.** 2009. An atlas of chaperone–protein interactions in *Saccharomyces cerevisiae*: implications to protein folding pathways in the cell. Mol. Syst. Biol. **5:**275.

71. **Groom K. R. , H. C. Heyman, M. C. Steffen, L. Hawkins, and N. C. Martin.** 1998. Kluyveromyces lactis SEF1 and its Saccharomyces cerevisiae homologue bypass the unknown essential function, but not the mitochondrial RNase P function, of the S. cerevisiae RPM2 gene. Yeast **14:**77-87.

72. **Guarente L.** 1983. Yeast promoters and lacZ fusions designed to study expression of cloned genes in yeast. Methods Enzymol. **101:**181-91.

73. **Guarente L., and M. Ptashne** 1981. Fusion of Escherichia coli lacZ to the cytochrome c gene of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U S A. **78:**2199-203.

74. **Guarente L., R. Yocum, and P. Gifford.** 1982. A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. Proc. Natl. Acad. Sci. U S A **79:**7410-4.

75. **Ha N., K. Hellauer, and B. Turcotte.** 1996. Mutations in target DNA elements of yeast HAP1 modulate its transcriptional activity without affecting DNA binding. Nucleic Acids Res. **24:**1453-9.

76. **Hai T., and T. Curran.** 1991. Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proc. Natl. Acad. Sci. U S A **88:**3720-4.

77. **Hall D.A. , H. Zhu, X. Zhu, T. Royce, M. Gerstein, and M. Snyder.** 2004. Regulation of gene expression by a metabolic enzyme. Science **306:**482-4.

78. **Harbison C. T. , G. D., T. I. Lee , N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J. B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel, and R. A. Young.** 2004. Transcriptional regulatory code of a eukaryotic genome. Nature **431:**99-104.

79. **Harrison, S. C., and A. K. Aggarwal.** 1990. DNA recognition by proteins with the helix-turn-helix motif. Annu. Rev. Biochem. **59:**933-69.

80. **Hellauer, K., M. H. Rochon, and B. Turcotte.** 1996. A novel DNA binding motif for yeast zinc cluster proteins: the Leu3p and Pdr3p transcriptional activators recognize everted repeats. Mol Cell Biol **16:**6096-102.

81. **Herrgård M. J., M. W. Covert, and B. Ø. Palsson.** 2004. Reconstruction of microbial transcriptional regulatory networks. Curr. Opin. Biotechnol. **15:**70-7.

82. **Hertz G. Z. , and G. D. Stormo** 1999. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics **15:**563-77.

83. **Ho S. W., G. Jona, C. T. Chen, M. Johnston, and M. Snyder.** 2006. Linking DNA-binding proteins to their recognition sequences by using protein microarrays. Proc. Natl. Acad. Sci. U S A **103:**9940-5.

84. **Hope I. A., and K. Struhl.** 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell **46:**885-94.

85. **Horak C. E., and M. Snyder.** 2002. ChIP-chip: a genomic approach for identifying transcription factor binding sites. Methods in Enzymology **350:**468-83.

86. **Hudson M. E., and M. Snyder.** 2006. High-throughput methods of regulatory element discovery. Biotechniques **41:**673, 675, 677 passim.

87. **Hughes J. R., J. F. Cheng, N. Ventress, S. Prabhakar, K. Clark, E. Anguita, M. De Gobbi, P. de Jong, E. Rubin, and D. R. Higgs.** 2005. Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. Proc. Natl. Acad. Sci. U S A **102:**9830-5.

88. **Hughes T. R., M. D. Robinson, N. Mitsakakis, and M. Johnston.** 2004. The promise of functional genomics: completing the encyclopedia of a cell. Curr. Opin. Microbiol. **7:**546-54.

89. **Imanishi T., T. Itoh, Y. Suzuki, C. O'Donovan, S. Fukuchi, K. O. Koyanagi, R. A. Barrero, T. Tamura, Y. Yamaguchi-Kabata, M. Tanino,**

K. Yura, S. Miyazaki, K. Ikeo, K. Homma, A. Kasprzyk, T. Nishikawa, M. Hirakawa, J. Thierry-Mieg, D. Thierry-Mieg, J. Ashurst, L. Jia, M. Nakao, M. A. Thomas, N. Mulder, Y. Karavidopoulou, L. Jin, S. Kim, T. Yasuda, B. Lenhard, E. Eveno, Y. Suzuki, C. Yamasaki, J. Takeda, C. Gough, P. Hilton, Y. Fujii, H. Sakai, S. Tanaka, C. Amid, M. Bellgard, Mde. F. Bonaldo, H. Bono, S. K. Bromberg, A. J. Brookes, E. Bruford, P. Carninci, C. Chelala, C. Couillault, S. J. de Souza, M. A. Debily, M. D. Devignes, I. Dubchak, T. Endo, A. Estreicher, E. Eyras, K. Fukami-Kobayashi, G. R. Gopinath, E. Graudens, Y. Hahn, M. Han, Z. G. Han, K. Hanada, H. Hanaoka, E. Harada, K. Hashimoto, U. Hinz, M. Hirai, T. Hishiki, I. Hopkinson, S. Imbeaud, H. Inoko, A. Kanapin, Y. Kaneko, T. Kasukawa, J. Kelso, P. Kersey, R. Kikuno, K. Kimura, B. Korn, V. Kuryshev, I. Makalowska, T. Makino, S. Mano, R. Mariage-Samson, J. Mashima, H. Matsuda, H. W. Mewes, S. Minoshima, K. Nagai, H. Nagasaki, N. Nagata, R. Nigam, O. Ogasawara, O. Ohara, M. Ohtsubo, N. Okada, T. Okido, S. Oota, M. Ota, T. Ota, T. Otsuki, D. Piatier-Tonneau, A. Poustka, S. X. Ren, N. Saitou, K. Sakai, S. Sakamoto, R. Sakate, I. Schupp, F. Servant, S. Sherry, R. Shiba, N. Shimizu, M. Shimoyama, A. J. Simpson, B. Soares, C. Steward, M. Suwa, M. Suzuki, A. Takahashi, G. Tamiya, H. Tanaka, T. Taylor, J. D. Terwilliger, P. Unneberg, V. Veeramachaneni, S. Watanabe, L. Wilming, N. Yasuda, H. S Yoo, M. Stodolsky, W. Makalowski, M. Go, K. Nakai, T. Takagi, M. Kanehisa, Y. Sakaki, J. Quackenbush, Y. Okazaki, Y. Hayashizaki, W.

Hide, R. Chakraborty, K. Nishikawa, H. Sugawara, Y. Tateno, Z. Chen, M. Oishi, P. Tonellato, R. Apweiler, K. Okubo, L. Wagner, S. Wiemann, R. L. Strausberg, T. Isogai, C. Auffray, N. Nomura, T. Gojobori, and S. Sugano. 2004. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. Plos. Biol. **2:**e162.

90. **Inai T., M. Yukawa, and E. Tsuchiya.** 2007. Interplay between chromatin and trans-acting factors on the IME2 promoter upon induction of the gene at the onset of meiosis. Mol. Cell Biol. **27:**1254-63.

91. **Indu Jagannathan, and J. J. Hayes** 2008. Hydroxyl Radical Footprinting of Protein–DNA Complexes. Methods in Molecular Biology **543:**57-71.

92. **Ito T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki.** 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. U S A **98:**4569-74.

93. **Iyer V. R., C. E. Horak, C. S. Scafe, D. Botstein, M. Snyder, and P. O. Brown.** 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature **409:**533-8.

94. **Jackman J. E., E. Grayhack, and E. M. Phizicky.** 2008. The use of Saccharomyces cerevisiae proteomic libraries to identify RNA-modifying proteins. Methods Molecular Biology **488**: 383-93.

95. **James P., J. Halladay, and E. A. Craig.** 1996. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics **144:**1425-36.

96. **Jelinsky S. A., P. Estep, G. M. Church, and L. D. Samson.** 2000. Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links base excision repair with proteasomes. Mol. Cell Biol. **20:**8157-8167.

97. **Johnson P. F., and S. L. McKnight.** 1989. Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. **58:**799-839.

98. **Johnston M, and J. Dover** 1987. Mutations that inactivate a yeast transcriptional regulatory protein cluster in an evolutionarily conserved DNA binding domain. Proc. Natl. Acad. Sci U S A **84:**2401-5.

99. **Johnston M., and J. Dover.** 1988. Mutational analysis of the GAL4-encoded transcriptional activator protein of Saccharomyces cerevisiae. Genetics **120:**63-74.

100. **Jordan S. R., and C. O. Pabo.** 1988. Structure of the lambda complex at 2.5 A resolution: details of the repressor-operator interactions. Science **242:**893-9.

101. **Kadonaga J. T.** 2004. Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. Cell **116:**247-57.

102. **Kahl B.F., and M. Paule.** 2008. The Use of Diethyl Pyrocarbonate and Potassium Permanganate as Probes for Strand Separation and Structural Distortions in DNA. Methods in Molecular Biology **453:**73-85.

103. **Karpichev I.V, Y. Luo, R. C. Marians, and G. M. Small.** 1997. A complex containing two transcription factors regulates peroxisome proliferation and the coordinate induction of beta-oxidation enzymes in Saccharomyces cerevisiae. Mol. Cell Biol. **17:**69-80.

104. **Karpichev I. V., and G. M. Small.** 1998. Global regulatory functions of Oaf1p and Pip2p (Oaf2p), transcription factors that regulate genes encoding peroxisomal proteins in Saccharomyces cerevisiae. Mol. Cell. Biol. **18:**6560-70.

105. **Keleher C. A., S. Passmore, and A. D. Johnson.** 1989. Yeast repressor alpha 2 binds to its operator cooperatively with yeast protein Mcm1. Mol. Cell Biol. **9:**5228-30.

106. **Kellis M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander.** 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423:**241-54.

107. **Kim T. H., L. O. Barrera, M. Zheng, C. Qu, M. A. Singer, T. A. Richmond, Y. Wu, R. D. Green, and B. Ren.** 2005. A high-resolution map of active promoters in the human genome. Nature **436:**876-80.

108. **Kissinger C. R., B. S. Liu, E. Martin-Blanco, T. B. Kornberg, and C. O. Pabo.** 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. Cell **63:**579-90.

109. **Ko L. J., and J. D. Engel.** 1993. DNA-binding specificities of the GATA transcription factor family. Mol. Cell Biol. **13:**4011-22.

110. **Koonin E. V., R. L. Tatusov, and K. E. Rudd.** 1995. Sequence similarity analysis of Escherichia coli proteins: functional and evolutionary implications. Proc. Natl. Acad. Sci. U S A **92:**11921-5.

111. **Kornberg, R. D.** 2001. The eukaryotic gene transcription machinery. Biol. Chem. **382:**1103-7.

112. **Kummerfeld S. K., and S. A. Teichmann** 2006. DBD: a transcription factor prediction database. Nucleic Acids Res. **34:**D74-D81.

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

114. **Labbé S., J. F. Harrisson, and C. Séguin.** 2009. Identification of sequence-specific DNA-binding proteins by southwestern blotting. Methods Molecular Biology **543:**151-61.

115. **Laity J. H., B. M. Lee, and P. E. Wright.** 2001. Zinc finger proteins: new insights into structural and functional diversity. Curr Opin Struct Biol **11:**39-46.

116. **Landschulz W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240:**1759-64.

117. **Laughon A., and M. P. Scott.** 1984. Sequence of a Drosophila segmentation gene: protein structure homology with DNA-binding proteins. Nature **310:**25-31.

118. **Lawrence C. E., S. Altschul, M. S. Boguski, J. S. Liu, A. F. Neuwald, and J. C. Wootton.** 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science **262:**208-14.

119. **Leblanc B., and T. Moss.** 2008. DNA-Protein Interactions: Principles and Protocols, Third Edition, vol. 543.

120. **Leblanc B., and T. Moss.** 2009. DNase I Footprinting. methods in Molecular Biology **543:**37-47.

121. **Lechner J.** 1994. A zinc finger protein, essential for chromosome segregation, constitutes a putative DNA binding subunit of the Saccharomyces cerevisiae kinetochore complex, Cbf3. Embo J **13:**5203-11.

122. **Lee T. I., N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young.** 2002. Transcriptional regulatory networks in Saccharomyces cerevisiae. Science **298:**799-804.

123. **Levine M, and R. Tjian.** 2003. Transcription regulation and animal diversity. Nature **424:**147-51.

124. **Li B., and J. C. Reese.** 2001. Ssn6-Tup1 regulates RNR3 by positioning nucleosomes and affecting the chromatin structure at the upstream repression sequence. J Biol Chem **276:**33788-97.

125. **Li J. J., and I. Herskowitz.** 1993. Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. Science **262:**1870-4.

126. **Liang S. D., R. Marmorstein, S. C. Harrison, and M. Ptashine.** 1996. DNA sequence preferences of GAL4 and PPR1: how a subset of Zn2 Cys6 binuclear cluster proteins recognizes DNA. Mol. Cell Biol. **16:**3773-80.

127. **Liu X., D. M. Noll, J. D. Lieb, and N. D. Clarke.** 2005. DIP-chip: rapid and accurate determination of DNA-binding specificity. Genome Res. **15:**421-7.

128. **Lohr D., P. Venkov, and J. Zlatanova.** 1995. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J **9:**777-87.

129. **Longtine M. S., A. 3rd McKenzie, D. J. Demarini, N. G. Shah, A. Wach, A. Brachat, P. Philippsen, and J. R. Pringle.** 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast **14:**953-61.

130. **Luo Y., S. Vijaychander, J. Stile, and L. Zhu.** 1996. Cloning and analysis of DNA-binding proteins by yeast one-hybrid and one-two-hybrid systems. Biotechniques **20:**564-8.

131. **Ma J., and M. Ptashne.** 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. Cell **48:**847-53.

132. **Ma J., and M. Ptashne.** 1987. A new class of yeast transcriptional activators. Cell **51:**113-9.

133. **MacIsaac K. D., T. Wang, D. B. Gordon, D. K. Gifford, G. D. Stormo, and E. Fraenkel.** 2006. An improved map of conserved regulatory sites for Saccharomyces cerevisiae. BMC BioInformatics **7:**113.

134. **Mackay J. P., and M. Crossley.** 1998. Zinc fingers are sticking together. Trends Biochem. Sci. **23:**1-4.

135. **MacPherson S., M. Larochelle, and B. Turcotte.** 2006. A fungal family of transcriptional regulators: the zinc cluster proteins. Microbiol. Mol. Biol. Rev. **70:**583-604.

136. **Mamane Y., K. Hellauer, M. H. Rochon, and B. Turcotte.** 1998. A linker region of the yeast zinc cluster protein leu3p specifies binding to everted repeat DNA. J. Biol. Chem. **273:**18556-61.

137. **Man T. K., G. D. Stormo.** 2001. Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay. Nucleic Acids Res. **29:**2471-8.

138. **Manfield I., and P. Stockley.** 2008. Ethylation Interference Footprinting of DNA–Protein Complexes. Methods in Molecular Biology **453:**105-120.

139. **Marmorstein R., M. Carey, M. Ptashne, and S. C. Harrison.** 1992. DNA recognition by GAL4: structure of a protein-DNA complex. Nature **356:**408-14.

140. **Marmorstein R, and S. C. Harrison.** 1994. Crystal structure of a PPR1-DNA complex: DNA recognition by proteins containing a Zn2Cys6 binuclear cluster. Genes Dev. **8:**2504-12.

141. **Massie C. E., and I. G. Mills.** 2008. ChIPping away at gene regulation. EMBO Rep. **9:**337-43.

142. **McCue L., W. Thompson, C. Carmack, M. P. Ryan, J. S. Liu, V. Derbyshire, and C. E. Lawrence.** 2001. Phylogenetic footprinting of transcription factor binding sites in proteobacterial genomes. Nucleic Acids Res. **29:**774-82.

143. **McEwen G. K., A. Woolfe, D. Goode, T. Vavouri, H. Callaway, and G. Elgar.** 2006. Ancient duplicated conserved noncoding elements in vertebrates: a genomic and functional analysis. Genome Res. **16:**451-65.

144. **McKay D. B., and T. A. Steitz.** 1981. Structure of catabolite gene activator protein at 2.9 A resolution suggests binding to left-handed B-DNA. Nature **290:**744-9.

145. **McNally J. G., W. G. Muller, D. Walker, R. Wolford, and G. L. Hager.** 2000. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. Science **287:**1262-5.

146. **Melcher K., B. Sharma, W. V. Ding WV, and M. Nolden.** 2000. Zero background yeast reporter plasmids. Gene **247:**53-61.

147. **Meng X., M. H. Brodsky, and S. A. Wolfe.** 2005. A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. Nature Biotechnology **23:**988-94.

148. **Meng X., and S. A. Wolfe.** 2006. Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. Nature Protocols **1:**30-45.

149. **Michael S. F., V. J. Kilfoil, M. H. Schmidt, B. T. Amann, and J. M. Berg.** 1992. Metal binding and folding properties of a minimalist Cys2His2 zinc finger peptide. Proc. Natl. Acad. Sci. U S A **89:**4796-800.

150. **Miller.** 1972. *Experiments in Molecular Genetics,Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY***:**352-355.

151. **Miller J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. **4:**1609-14.

152. **Mirny L. A., and M. S. Gelfand.** 2002. Structural analysis of conserved base pairs in protein-DNA complexes. Nucleic Acids Res. **30:**1704-11.

153. **Mittler G., F. Butter, and M. Mann.** 2009. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Res. **19:**284-93.

154. **Mueller P. R., and B. Wold.** 1989. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science **246:**780-6.

155. **Mukherjee S., M. F. Berger, G. Jona, X. S. Wang, D. Muzzey, M. Snyder, R. A. Young, and M. L. Bulyk.** 2004. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. Nat. Genet. **36:**1331-9.

156. **Mukherjee S., M. F. Berger, G. Jona, X. S. Wang, D. Muzzey, M. Snyder, R. A. Young, and M. L. Bulyk.** 2004. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. Nature Genetics **36:**1331-9.

157. **Murre, C., P. S. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell **56:**777-83.

158. **Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell **58:**537-44.

159. **Nardelli, J., T. Gibson, and P. Charnay.** 1992. Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis. Nucleic Acids Res. **20:**4137-44.

160. **Nardelli, J., T. J. Gibson, C. Vesque, and P. Charnay.** 1991. Base sequence discrimination by zinc-finger DNA-binding domains. Nature **349:**175-8.

161. **Newburger D.E., and M. L. Bulyk.** 2009. UniPROBE: an online database of protein binding microarray data on protein–DNA interactions. Nucleic Acids Res. **37:**D77-82.

162. **Nielsen P. E.** 2008. Uranyl Photofootprinting. Methods in Molecular Biology **453**.

163. **O'Shea, E. K., J. D. Klemm, P. S. Kim, and T. Alber.** 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science **254:**539-44.

164. **Ohlendorf D. H., W. F. Anderson, R. G. Fisher, Y. Takeda, and B. W. Matthews.** 1982. The molecular basis of DNA-protein recognition inferred from the structure of cro repressor. Nature **298:**718-23.

165. **Oliphant A. R., C. J. Brandl, and K. Struhl.** 1989. Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. Mol. Cell Biol. **9:**2944-2949.

166. **Orian A., B. van Steensel, J. Delrow, H. J. Bussemaker, L. Li, T. Sawado, E. Williams, L. W. Loo, S. M. Cowley, C. Yost, S. Pierce, B. A. Edgar, S. M. Parkhurst, and R. N. Eisenman.** 2003. Genomic binding by the

Drosophila Myc, Max, Mad/Mnt transcription factor network. Genes Dev. **17:**1101-14.

167. **Ostling J., M. Carlberg, and H. Ronne.** 1996. Functional domains in the Mig1 repressor. Mol. Cell Biol. **16:**753-61.

168. **Pabo C. O., and M. Lewis.** 1982. The operator-binding domain of lambda repressor: structure and DNA recognition. Nature **298:**443-7.

169. **Pabo C. O., and R. T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61:**1053-95.

170. **Pan T., and J. E. Coleman.** 1990. GAL4 transcription factor is not a "zinc finger" but forms a Zn(II)2Cys6 binuclear cluster. Proc. Natl. Acad. Sci U S A **87:**2077-81.

171. **Pan T., and J. E. Coleman.** 1991. Sequential assignments of the 1H NMR resonances of Zn(II)2 and 113Cd(II)2 derivatives of the DNA-binding domain of the GAL4 transcription factor reveal a novel structural motif for specific DNA recognition. Biochemistry **30:**4212-22.

172. **Papoulas O.** 2001. Rapid separation of protein-bound DNA from free DNA using nitrocellulose filters. Curr Protoc Mol Biol. **Chapter 12:**Unit 12.8.

173. **Parrish J. R., K. D. Gulyas, and R. L. Jr Finley.** 2006. Yeast two-hybrid contributions to interactome mapping. Curr Opin Biotechnol. **17:**387-93.

174. **Patrizia Spitalny, and M. Thomm.** 2008. Exonuclease III Footprinting on Immobilized DNA Templates. Methods in Molecular Biology **543:**49-56.

175. **Pedersen J. T., and J. Moult.** 1996. Genetic algorithms for protein structure prediction. Curr Opin Struct Biol **6:**227-31.

176. **Phizicky E. M., and E. Grayhack.** 2006. Proteome-scale analysis of biochemical activity. Crit. Rev. Biochem. Mol. Biol. **41:**315-27.

177. **Pilpel Y., Sudarsanam, P., and Church, G.M.** 2001. Identifying regulatory networks by combinatorial analysis of promoter elements. Nat. Genet. **29:**153-159.

178. **Pir P., K. Ulgen, A. Hayes, Z. Ilsen Onsan Z, B. Kirdar, and S. G. Oliver.** 2006. Annotation of unknown yeast ORFs by correlation analysis of microarray data and extensive literature searches.Yeast **23:**553-71.

179. **Polish J., J. H. Kim, and M. Johnston.** 2005. How the Rgt1 Transcription Factor of Saccharomyces cerevisiae is regulated by glucose. Genetics **169:**583-594.

180. **Popova Y., P. Thayumanavan, E. Lonati, M. Agrochão, and J. M Thevelein.** 2010. Transport and signaling through the phosphate-binding site of the yeast Pho84 phosphate transceptor. Proc. Natl. Acad. Sci. U S A. **107:**2890-5.

181. **Post-Beittenmiller M. A., R. W. Hamilton, and J. E. Hopper.** 1984. Regulation of basal and induced levels of the MEL1 transcript in Saccharomyces cerevisiae. Mol. Cell Biol. **4:**1238-45.

182. **Qian J., J. Lin, N. M. Luscombe, H. Yu, and M. Gerstein.** 2003. Prediction of regulatory networks: genome-wide identification of transcription factor targets from gene expression data. Bioinformatics **19:**1917-26.

183. **Qian Y. Q., M. Billeter, G. Otting, M. Muller, W. J. Gehring, and K. Wuthrich.** 1989. The structure of the Antennapedia homeodomain

determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell **59:**573-80.

184. **Qiu P.** 2003. Recent advances in computational promoter analysis in understanding the transcriptional regulatory network. Biochem. Biophys. Res. Commun. **309:**495-501.

185. **Rasmussen R., D. Benvegnu, E. K. O'Shea, P. S. Kim, and T. Alber.** 1991. X-ray scattering indicates that the leucine zipper is a coiled coil. Proc. Natl. Acad. Sci. U S A **88:**561-4.

186. **Rauscher F. J., 3rd, D. R. Cohen, T. Curran, T. J. Bos, P. K. Vogt, D. Bohmann, R. Tjian, and B. R. Franza, Jr.** 1988. Fos-associated protein p39 is the product of the jun proto-oncogene. Science **240:**1010-6.

187. **Reece R. J., and M. Ptashne.**1993. Determinants of binding-site specificity among yeast C6 zinc cluster proteins. Science **261:**909-11.

188. **Reece-Hoyes J. S., B. Deplancke, J. Shingles, C. A. Grove, I. A. Hope, and A. J. Walhout.** 2005. A compendium of Caenorhabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biol. **6:**R110.

189. **Ren B., F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell, and R. A. Young.** 2000. Genome-wide location and function of DNA binding proteins. Science **290:**2306-9.

190. **Ren, B., F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J.**

Wilson, S. P. Bell, and R. A. Young. 2000. Genome-wide location and function of DNA binding proteins. Science **290:**2306-9.

191. **Roberts G. G., and A. P. Hudson** 2006. Transcriptome profiling of Saccharomyces cerevisiae during a transition from fermentative to glycerol-based respiratory growth reveals extensive metabolic and structural remodeling. Mol. Genet. Genomics **276:**170-86.

192. **Robertson G., M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney, N. Thiessen, O. L. Griffith, A. He, M. Marra, M. Snyder, and S. Jones.** 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nature Methods **4:**651-7.

193. **Rottensteiner H., A. J. Kal, B. Hamilton, H. Ruis, and H. F. Tabak.** 1997. A heterodimer of the Zn2Cys6 transcription factors Pip2p and Oaf1p controls induction of genes encoding peroxisomal proteins in Saccharomyces cerevisiae. Eur. J. Biochem. **247:**776-83.

194. **Roulet E., S. Busso, A. A. Camargo, A. J. Simpson, N. Mermod, P. Bucher.** 2002. High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. Nat. Biotechnol. **20:**831-5.

195. **Rubin G. M., M. D. Yandell, J. R. Wortman, G. L. Gabor Miklos, C. R. Nelson, I. K. Hariharan, M. E. Fortini, P. W. Li, R. Apweiler, W. Fleischmann, J. M. Cherry, S. Henikoff, M. P. Skupski, S. Misra, M. Ashburner, E. Birney, M. S. Boguski, T. Brody, P. Brokstein, S. E. Celniker, S. A. Chervitz, D. Coates, A. Cravchik, A. Gabrielian, R. F.**

Galle, W. M. Gelbart, R. A. George, L. S. Goldstein, F. Gong, P. Guan, N. L. Harris, B. A. Hay, R. A. Hoskins, J. Li, Z. Li, R. O. Hynes, S. J. Jones, P. M. Kuehl, B. Lemaitre, J. T. Littleton, D. K. Morrison, C. Mungall, P. H. O'Farrell, O. K. Pickeral, C. Shue, L. B. Vosshall, J. Zhang, Q. Zhao, X. H. Zheng, and S. Lewis. 2000. Comparative genomics of the eukaryotes. Science **287:**2204-15.

196. **Sabina J., and M. Johnston.** 2009. Asymmetric signal transduction through paralogs that comprise a genetic switch for sugar sensing in Saccharomyces cerevisiae. J. Biol. Chem. **284:**29635-43.

197. **Sauer R. T., D. L. Smith, and A. D. Johnson.** 1988. Flexibility of the yeast alpha 2 repressor enables it to occupy the ends of its operator, leaving the center free. Genes Dev. **2:**807-16.

198. **Schjerling P., and S. Holmberg.** 1996. Comparative amino acid sequence analysis of the C6 zinc cluster family of transcriptional regulators. Nucleic Acids Res. **24:**4599-607.

199. **Schmid M., T. Drussel, and U. K. Laemmli.** 2004. ChIC and ChEC; genomic mapping of chromatin proteins. Mol. Cell **16:**147-57.

200. **Schneider T. D. and R. M. Stephens.** 1990. Sequence logos: A new way to display consensus sequences. Nucleic Acids Res. **18:**6097-6100.

201. **Schüller H. J., A. Hahn, F. Tröster, A. Schütz, and E. Schweizer.** 1992. Coordinate genetic control of yeast fatty acid synthase genes FAS1 and FAS2 by an upstream activation site common to genes involved in membrane lipid biosynthesis. EMBO J. **11:**107-14.

202. **Scott M. P., J. W. Tamkun, and G. W. Hartzell, 3rd.** 1989. The structure and function of the homeodomain. Biochim Biophys. Acta. **989:**25-48.

203. **Seeman N. C., J. M. Rosenberg, and A. Rich.** 1976. Sequence-specific recognition of double helical nucleic acids by proteins. Proc Natl Acad Sci U S A **73:**804-8.

204. **Setzer D., D. Schulman, C. Gunther, and M. Bumbulis.** 2008. Use of a Reporter Gene Assay in Yeast for Genetic Analysis of DNA–Protein Interactions. Methods in Molecular Biology **543:**219-241.

205. **Shen H., Y. Zhang, H. Ding, X. Wang, L. Chen, H. Jiang, and X. Shen.** 2008. Farnesoid X receptor induces GLUT4 expression through FXR response element in the GLUT4 promoter. Cell Phsiol. Biochem. **22:**1-14.

206. **Shepherd J. C., W. McGinnis, A. E. Carrasco, E. M. De Robertis, and W. J. Gehring.** 1984. Fly and frog homoeo domains show homologies with yeast mating type regulatory proteins. Nature **310:**70-1.

207. **Siddharthan R.** 2008. PhyloGibbs-MP: module prediction and discriminative motif-finding by Gibbs sampling. PLos Comput Biol **4:**e1000156.

208. **Sikorski R. S., P. Hieter** 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics **122:**19-27.

209. **Simon M., G. Adam, W. Rapatz, W. Spevak and H. Ruis.** 1991. The Saccharomyces cerevisiae ADR1 gene is a positive regulator of transcription of genes encoding peroxisomal proteins. Mol. Cell Biol. **11:**699-704.

210. **Stansfield HE, K. B., Lybrand KE, Jamieson ER.** 2009. Identifying protein interactions with metal-modified DNA using microarray technology. J Biol Inorg Chem. **14:**193-9.

211. **Stavreva D. A., and J. G. McNally.** 2004. Fluorescence recovery after photobleaching (FRAP) methods for visualizing protein dynamics in living mammalian cell nuclei. Methods Enzymol. **375:**443-55.

212. **Stein, L. D., Z. Bao, D. Blasiar, T. Blumenthal, M. R. Brent, N. Chen, A. Chinwalla, L. Clarke, C. Clee, A. Coghlan, A. Coulson, P. D'Eustachio, D. H. Fitch, L. A. Fulton, R. E. Fulton, S. Griffiths-Jones, T. W. Harris, L. W. Hillier, R. Kamath, P. E. Kuwabara, E. R. Mardis, M. A. Marra, T. L. Miner, P. Minx, J. C. Mullikin, R. W. Plumb, J. Rogers, J. E. Schein, M. Sohrmann, J. Spieth, J. E. Stajich, C. Wei, D. Willey, R. K. Wilson, R. Durbin, and R. H. Waterston.** 2003. The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol **1:**E45.

213. **Steitz, T. A., D. H. Ohlendorf, D. B. McKay, W. F. Anderson, and B. W. Matthews.** 1982. Structural similarity in the DNA-binding domains of catabolite gene activator and cro repressor proteins. Proc. Natl. Acad. Sci. U S A **79:**3097-100.

214. **Stoevesandt O., M. J. Taussig, and M. He.** 2009. Protein microarrays: high-throughput tools for proteomics. Expert Rev. Proteomics **6:**145-57.

215. **Stormo G. D., and D. S. Fields** 1998. Specificity, free energy and information content in protein-DNA interactions. Trends Biochem. Sci. **23:**109-13.

216. **Stribinskis V., H. C. Heyman, S. R. Ellis, M. C. Steffen, N. C. Martin.** 2005. **Rpm2p, a component of yeast mitochondrial RNase P, acts as a transcriptional activator in the nucleus.** Mol. Cell Biol. **25:**6546-58.

217. **Swinnen E., V. Wanke, J. Roosen, B. Smets, F. Dubouloz, I. Pedruzzi, E. Cameroni, C. De Virgilio, and J. Winderickx.** 2006. Rim15 and the crossroads of nutrient signalling pathways in Saccharomyces cerevisiae. Cell Div. 2006 Apr 3;1:3.

218. **Talanian R. V., C. J. McKnight, and P. S. Kim.** 1990. Sequence-specific DNA binding by a short peptide dimer. Science **249:**769-71.

219. **Taunton J., C. A. Hassig, and S. L. Schreiber.** 1996. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science **272:**408-11.

220. **Thevelein J. M., and J. H. de Winde.** 1999. Novel sensing mechanisms and targets for the cAMP-protein kinase A pathway in the yeast Saccharomyces cerevisiae. Mol. Microbiol. **33:**904-18.

221. **Todd R. B., and A. Andrianopoulos.** 1997. Evolution of a fungal regulatory gene family: the Zn(II)2Cys6 binuclear cluster DNA binding motif. Fungal Genet Biol **21:**388-405.

222. **Tompa M., N. Li, T. L. Bailey, G. M. Church, B. De Moor, E. Eskin, A. V. Favorov, M. C. Frith, Y. Fu, W. J. Kent, V. J. Makeev, A. A. Mironov, W.**

S. Noble, G. Pavesi, G. Pesole, M. Régnier, N. Simonis, S. Sinha, G. Thijs, J. van Helden, M. Vandenbogaert, Z. Weng, C. Workman, C. Ye, and Z. Zhu. 2005. Assessing computational tools for the discovery of transcription factor binding sites. Nature Biotechnology **23:**137-44.

223. **Toyn J. H., P. L. Gunyuzlu, W. H. White, L. A. Thompson, and G. F. Hollis.** 2000. A counterselection for the tryptophan pathway in yeast: 5-fluoroanthranilic acid resistance. Yeast. **16:**553-60.

224. **Tsang, A., J. Visvader, C. Turner, Y. Fujiwara, C. Yu, M. J. Weiss, M. Crossley, and S. H. Orkin.** 1997. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. Cell **90:**109-19.

225. **Turcotte B., X. B. Liang, F. Robert, and N. Soontorngun.** 2010. Transcriptional regulation of nonfermentable carbon utilization in budding yeast. FEMS Yeast Res. **10:**2-13.

226. **Uetz P., L. Giots, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg.** 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature **403:**623-7.

227. **Urnov, F. D.** 2002. A feel for the template: zinc finger protein transcription factors and chromatin. Biochem. Cell Biol. **80:**321-33.

228. **van Bakel H., F. J. van Werven, M. Radonjic, M. O. Brok MO, D. van Leenen, F. C. Holstege, and H. T. Timmers.** 2008. Improved genome-wide localization by ChIP-chip using double-round T7 RNA polymerase-based amplification. Nucleic Acids Res. **36:**e21.

229. **van Steensel B., J. Delrow, and S. Henikoff.** 2001. Chromatin profiling using targeted DNA adenine methyltransferase. Nature Genetics **27:**304-8.

230. **van Steensel B., and S. Henikoff.** 2000. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nature Biotechnology **18:**424-8.

231. **Vavouri T., and G. Elgar.** 2005. Prediction of cis-regulatory elements using binding site matrices--the successes, the failures and the reasons for both. Curr. Opin. Genet. Dev. **15:**395-402.

232. **Venkatesh B., and W. H. Yap.** 2005. Comparative genomics using fugu: a tool for the identification of conserved vertebrate cis-regulatory elements. Bioessays **27:**100-7.

233. **Vermeulen W., and A. B. Houtsmuller.** 2002. The transcription cycle in vivo. A blind watchmaker at work. Mol. Cell.\ **10:**1264-6.

234. **Vidal M., and R. F. Gaber.** 1991. RPD3 encodes a second factor required to achieve maximum positive and negative transcriptional states in Saccharomyces cerevisiae. Mol. Cell Biol. **11:**6317-27.

235. **Vik A., and Rine J.** 2001. Upc2p and Ecm22p, dual regulators of sterol biosynthesis in Saccharomyces cerevisiae. Mol. Cell Biol. **21:**6395–6405.

236. **Vinson, C., A. Acharya, and E. J. Taparowsky.** 2006. Deciphering B-ZIP transcription factor interactions in vitro and in vivo. Biochim. Biophys. Acta. **1759:**4-12.

237. **Visser, A. E., P. J. Verschure, W. M. Gommans, H. J. Haisma, and M. G. Rots.** 2006. Step into the groove: engineered transcription factors as modulators of gene expression. Adv. Genet. **56:**131-61.

238. **Vlieghe D., A. Sandelin, P. J. De Bleser, K. Vleminckx, W. W. Wasserman, F. van Roy, and B. Lenhard.** 2006. A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. Nucleic Acids Res. **34:**D95-7.

239. **Vyas V. K., S. Kuchin, and M. Carlson.** 2001. Interaction of the repressors Nrg1 and Nrg2 with the Snf1 protein kinase in Saccharomyces cerevisiae. Genetics **158:**567-572.

240. **Walhout A. J.** 2006. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. Genome Res. **16:**1445-54.

241. **Walhout A. J., G. F. Temple, M. A. Brasch, J. L. Hartley, M. A. Lorson, S. van den Heuvel, and V. Vidal.** 2000. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. Methods Enzymol. **328:**575-92.

242. **Wang H., M. E. Heinz, S. D. Crosby, M. Johnston, and R. D. Mitra.** 2008. 'Calling Cards' method for high-throughput identification of targets of yeast DNA-binding proteins. Nature Protocols **3:**1569-77.

243. **Wang H., M. Johnston, and R. D. Mitra.** 2007. Calling cards for DNA-binding proteins. Genome Res. **17:**1202-9.

244. **Wang M. M., and R. R. Reed** 1993. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. Nature **364:**121-6.

245. **Wang, T., and G. D. Stormo.** 2003. Combining phylogenetic data with co-regulated genes to identify regulatory motifs. Bioinformatics **19:**2369-80.

246. **Watzele G., and W. Tanner.** 1989. Cloning of the glutamine:fructose-6-phosphate amidotransferase gene from yeast. Pheromonal regulation of its transcription. J. Biol. Chem. **264:**8753-8.

247. **Weinberger, C., S. M. Hollenberg, M. G. Rosenfeld, and R. M. Evans.** 1985. Domain structure of human glucocorticoid receptor and its relationship to the v-erb-A oncogene product. Nature **318:**670-2.

248. **Weinberger, C., C. C. Thompson, E. S. Ong, R. Lebo, D. J. Gruol, and R. M. Evans.** 1986. The c-erb-A gene encodes a thyroid hormone receptor. Nature **324:**641-6.

249. **Weintraub H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang, and A. Lassarand.** 1991. The myoD gene family: nodal point during specification of the muscle cell lineage. Science **251:**761-6.

250. **Weintraub H., V. J. Dwarki, I. Verma, R. Davis, S. Hollenberg, L. Snider, A. Lassar, and S. J. Tapscott.** 1991. Muscle-specific transcriptional activation by MyoD. Genes Dev. **5:**1377-86.

251. **Westendorf J. J.** 2006. Transcriptional co-repressors of Runx2. J. Cell Biochem. **98:**54-64.

252. **Wingender E., P. Dietze, H. Karas, and R. Knüppel.** 1996. TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res. **24:**238-41.

253. **Wolfe S. A., L. Nekludova, and C. O. Pabo.** 2000. DNA recognition by Cys2His2 zinc finger proteins. Annu. Rev. Biophys. Biomol. Struct. **29:**183-212.

254. **Wolfinger R. D., G. Gibson, E. D. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R. S. Paules.** 2001. Assessing gene significance from cDNA microarray expression data via mixed models. J. Comput. Biol. **8:**625-37.

255. **Wong I, T. M. Lohman** 1993. A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. Proc. Natl. Acad. Sci. U S A **90:**5428-32.

256. **Woolfe A., M. Goodson, D. K. Goode, P. Snell, G. K. McEwen, T. Vavouri, S. F. Smith, P. North, H. Callaway, K. Kelly, K. Walter, I. Abnizova, W. Gilks, Y. J. Edwards, J. E. Cooke, and G. Elgar.** 2005. Highly conserved non-coding sequences are associated with vertebrate development. PLos Biol. **3:**e7.

257. **Workman C. T., H. C. Mak, S. McCuine, J. B. Tagne, M. Agarwal, O. Ozier, T. J. Begley, L. D. Samson, and T. Ideker.** 2006. A systems approach to mapping DNA damage response pathways. Science **312:**1054-9.

258. **Workman, C. T., Y. Yin, D. L. Corcoran, T. Ideker, G. D. Stormo, and P. V. Benos.** 2005. enoLOGOS: a versatile web tool for energy normalized sequence logos. Nucleic Acids Res. **33:**W389-92.

259. **Wykoff D. D., A. H Rizvi, J. M. Raser, B. Margolin, and E. K. O'Shea.** 2007. Positive feedback regulates switching of phosphate transporters in S. cerevisiae. Mol Cell **27:**1005-13.

260. **Yao, J., K. M. Munson, W. W. Webb, and J. T. Lis.** 2006. Dynamics of heat shock factor association with native gene loci in living cells. Nature **442:**1050-3.

261. **Young E. T., K. M. Dombek, C. Tachibana and T. Ideker.** 2003. Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. J. Biol. Chem. **278:**26146–26158.

262. **Yu H., P. Braun, M.A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J. F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A. S. de Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore, C. Boone, M. Snyder, F. P. Roth, A. L. Barabasi, J. Tavernier, D. E. Hill, and M. Vidal.** 2008. High-quality binary protein interaction map of the yeast interactome network. Science **322:**104.

263. **Zeitlinger J., R. P. Zinzen, A. Stark, M. Kellis, H. Zhang, R. A. Young, and M. Levine.** 2007. Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. Genes Dev. **21:**385-90.

264. **Zhang L., K. Zhang, R. Prändl, and F. Schöffl.** 2004. Detecting DNA-binding of proteins in vivo by UV-crosslinking and immunoprecipitation. Biochem Biophys. Res. Commun. **322:**705-11.

265. **Zhu C., K. J. Beyers, R. P. McCord, Z. Shi, M.F. Berger, D.E. Newburger, K. Saulrieta, Z. Smith, M. V. Shah, M. Radhakrishnan, A. A. Philippakis, Y. Hu, F. De Masi, M. Pacek, A. Rolfs, T. Murthy, J. Labaer, and M. L. Bulyk.** 2009. High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res. **19:**556-66.

266. **Zhu H., M. L. Bulyk, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, and M. Snyder.** 2001. Global analysis of protein activities using proteome chips. Science **293:**2101-5.