



## Identification of Target Sites of the $\alpha 2$ -Mcm1 Repressor Complex in the Yeast Genome

Hualin Zhong, Ron McCord and Andrew K. Vershon

*Genome Res.* 1999 9: 1040-1047

Access the most recent version at doi:[10.1101/gr.9.11.1040](https://doi.org/10.1101/gr.9.11.1040)

---

**References** This article cites 47 articles, 22 of which can be accessed free at:  
<http://genome.cshlp.org/content/9/11/1040.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the words "LEARN MORE". On the right, there is a photograph of a woman wearing a red and white superhero cape and mask, with the Cellecta logo (a green molecular structure) and the word "CELLECTA" below it.

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

## Research

# Identification of Target Sites of the $\alpha 2$ -Mcm1 Repressor Complex in the Yeast Genome

Hualin Zhong, Ron McCord, and Andrew K. Vershon<sup>1</sup>

Waksman Institute and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854-8020 USA

The  $\alpha 2$  and Mcm1 proteins bind DNA as a heterotetramer to repress transcription of cell-type-specific genes in the yeast *Saccharomyces cerevisiae*. Based on the DNA sequence requirements for binding by the  $\alpha 2$ -Mcm1 complex, we have searched the yeast genome for all potential  $\alpha 2$ -Mcm1 binding sites. Genes adjacent to the sites were examined for expression in the different cell mating types. These sites were further analyzed by cloning the sequences into a heterologous promoter and assaying for  $\alpha 2$ -Mcm1-dependent repression in vivo and DNA-binding affinity in vitro. Fifty-nine potential binding sites were identified in the search. Thirty-seven sites are located within or downstream of coding region of the gene. None of the sites assayed from this group are functional repressor sites in vivo or bound by the  $\alpha 2$ -Mcm1 complex in vitro. Among the remaining 22 sites, six are in the promoters of known  $\alpha$ -specific genes and two other sites have an  $\alpha 2$ -Mcm1-dependent role in determining the direction of mating type switching. Among the remaining sequences, we have identified a functional site located in the promoter region of a previously uncharacterized gene, *SCYJL170C*. This site functions to repress transcription of a heterologous promoter and the  $\alpha 2$ -Mcm1 complex binds to the site in vitro. *SCYJL170C* is repressed by  $\alpha 2$ -Mcm1 in vivo and therefore using this method we have identified a new  $\alpha$ -specific gene, which we call *ASG7*.

Many eukaryotic transcription factors have been identified based upon their strong amino acid sequence conservation to known DNA-binding domains and regulatory proteins. Often, subsequent analysis of the temporal or spatial expression of these proteins suggests that they may be involved in the regulation of critical cellular or developmental processes. To understand their roles in developmental and transcriptional regulation, the identification of their target genes is essential.

One common method used to identify the target genes is to remove or alter the activity of the regulatory protein, either through mutations or by changing its expression, and then examine the effects of these changes. The downstream target genes can be identified through subtractive DNA libraries, differential display, or by DNA microarray techniques (DeRisi et al. 1996; Lashkari et al. 1997; Chu et al. 1998; Lutfiyya et al. 1998; Roth et al. 1998). This approach works well for organisms such as the yeast *Saccharomyces cerevisiae*, in which it is possible to construct strains with a mutation in these genes and assay the effects of mutants (DeRisi et al. 1997; Lutfiyya et al. 1998). In higher eukaryotes, however, it is difficult or impossible to use this approach to screen for downstream target genes because it is difficult to alter the activity of the regulatory proteins.

As an alternative approach to identifying the target genes of transcriptional regulatory proteins, the

binding sites for these proteins can be determined through in vitro site selection methods (Blackwell and Weintraub 1990). This information is then used to search the genomic databases for genes containing these sites. The lack of a complete sequence of the organism has made this approach limited in the past. The genomic DNA sequences of many prokaryotic organisms, as well as the eukaryotes *S. cerevisiae* and *Caenorhabditis elegans*, however, have been reported over the last few years (Fleischmann et al. 1995; Bult et al. 1996; Goffeau et al. 1996; Deckert et al. 1998). The genome sequence projects of several other model organisms, such as *Drosophila* and *Arabidopsis*, are well under way (Dickson 1998; Flanders et al. 1998; Venter et al. 1998). As progress on the genome projects continues, it becomes possible to search an entire genome for all potential binding sites of a given DNA-binding protein. In this paper we describe this type of search to identify all of the target genes in yeast that are regulated by the  $\alpha 2$  and Mcm1 proteins that form a complex to bind DNA and repress expression of cell-type-specific genes. These experiments served to identify a new cell-type-specific gene in yeast. In addition, our results suggest that there are no extraneous  $\alpha 2$ -Mcm1 target sites in the genome and therefore all functional sites are regulating cell-type-specific functions.

## RESULTS

### Identification of Potential $\alpha 2$ -Mcm1 Binding Sites

There are three different cell types in the yeast *S. cerevisiae*—haploid **a**,  $\alpha$ , and diploid **a**/ $\alpha$  cells—which dif-

<sup>1</sup>Corresponding author.  
E-MAIL [vershon@mbcl.rutgers.edu](mailto:vershon@mbcl.rutgers.edu); FAX (732) 445-5735.

fer in mating specificity and ability to sporulate. The differences in cell type are attributable to the expression of different cell-type-specific genes under the control of the *MAT* locus (Herskowitz 1988; Johnson 1992). The *MAT* $\alpha$  locus encodes  $\alpha 2$ , a homeodomain protein, which is required for transcriptional repression of **a**-cell-type-specific genes in  $\alpha$  cells (Strathern et al. 1981; Tatchell et al. 1981; Herskowitz 1988; Johnson 1992). The  $\alpha 2$  protein interacts with Mcm1, a MADS-box protein that is similar to the mammalian serum response factor, SRF (Keleher et al. 1988; Passmore et al. 1988; Treisman 1995; Tan and Richmond 1998). The  $\alpha 2$  and Mcm1 proteins bind to a conserved, partially symmetric, 31-bp sequence located upstream of the transcription start site of **a**-specific genes (Johnson and Herskowitz 1985; Miller et al. 1985). The two proteins bind to the site as a heterotetramer, in which the Mcm1 dimer occupies the center region of the site and is flanked by  $\alpha 2$  monomers. Although both proteins bind to this site on their own, they bind cooperatively with more than 100-fold higher affinity (Keleher et al. 1988). The Mcm1 protein also functions to set the spacing and orientation for  $\alpha 2$  binding, which contributes to the specificity of the complex (Smith and Johnson 1992).

Previous studies have determined, in detail, the sequence requirements for the  $\alpha 2$  protein to bind DNA with higher affinity and cooperatively in complex with Mcm1. Although each position in the 31-bp  $\alpha 2$ -Mcm1 recognition site has a base-specific preference, highly conserved TGTA sequences in each half site are essential for  $\alpha 2$  binding and repression (Smith and Johnson 1994; Zhong and Vershon 1997). In the crystal structures of  $\alpha 2$  binding alone and in complex with Mcm1, residues in the  $\alpha 2$  homeodomain make specific contacts with these bases (Wolberger et al. 1991; Tan and Richmond 1998). In the center of the  $\alpha 2$ -Mcm1 site, the 5'-CCNNNNNNGG sequence has been showed to be required for Mcm1 binding (Treisman et al. 1989; Acton et al. 1997; West et al. 1997). The distance between the  $\alpha 2$  sites and the Mcm1 site is variable among the natural half sites, with a spacing of 4 or 5 bp. Sites with 4 or 5 bp spacing repress transcription equally well, but sites with 3 or 6 bp spacing fail to repress transcription and are not bound cooperatively by  $\alpha 2$  and Mcm1 in vitro (Mead et al. 1996). Although both  $\alpha 2$  and Mcm1 make base-specific contacts to these bases, mutagenesis results show that there is relaxed sequence specificity, with a preference for A or T bases at these positions (Acton et al. 1997; Zhong and Vershon 1997). Taking the data from all of these experiments together, a degenerate consensus sequence of TGTAN(A/T)<sub>3-4</sub>CCN<sub>6</sub>GG(A/T)<sub>3-4</sub>NTACA can be deduced for the  $\alpha 2$ -Mcm1 binding site.

The degenerate consensus sequence was used to search for all the potential  $\alpha 2$ -Mcm1 recognition sites

in the entire yeast genome using the PatMatch program provided by the *Saccharomyces* Genome Database. Allowing for a maximum of two mismatches in conserved bases, 59 sequences were obtained in the search (Table 1). All of the sites that perfectly match the search sequence are located in the promoter regions of known **a**-specific genes. Four sequences were identified which contain one mismatch with the search sequence. Two of these sites fall within the promoter regions between divergently transcribed open reading frames (ORFs) and therefore could potentially regulate both of these genes. Of the remaining 49 sites containing two mismatches with the degenerate consensus sequence, the majority of these are within ORFs. All of the previously identified  $\alpha 2$ -Mcm1 sites are within 250bp upstream of the translation start codon. It has been shown that the sites placed further upstream of the gene or downstream of the start of transcription are significantly weaker in their ability to repress transcription (Johnson and Herskowitz 1985; A.K.Vershon, unpubl.). Therefore, because many of the sites in this group are within the genes and contain multiple mismatches at positions important for binding they are unlikely to be functional  $\alpha 2$ -Mcm1 repressor sites.

#### Expression of Potential $\alpha$ -Specific Genes

To determine if any of the  $\alpha 2$ -Mcm1 sites identified in the search regulate transcription of nearby genes, the expression of these genes was compared in different cell types by RT-PCR and Northern assays. Genes that are regulated by the  $\alpha 2$ -Mcm1 complex are repressed in  $\alpha$  and diploid **a**/ $\alpha$  cells and expressed in **a** cells. An example of the assay is shown in Figure 1 and a summary of the results is shown in Table 2. In addition to the known **a**-specific genes, *STE2*, *STE6*, *MFA1*, *MFA2*, *BARI*, and *AGA2*, the only ORF that showed a similar pattern of expression in **a** versus  $\alpha$  cells was *YJL170C*. To confirm this result, the *YJL170C* coding region was used as a probe in a Northern blot of RNA prepared from **a**,  $\alpha$ , and diploid **a**/ $\alpha$  cells (Fig. 2). Our results show that *YJL170C* is expressed in **a** but not in  $\alpha$  or diploid **a**/ $\alpha$  cells and is therefore an **a**-cell-type-specific gene. This is the seventh **a**-specific gene that has been identified and we have named this ORF *ASG7*.

The start of the coding region of the *ASG7* gene is 397 bp away from the start of the *YJL169W* ORF. Because these genes are transcribed in the opposite direction, it is possible that they share promoter elements. In fact, the  $\alpha 2$ -Mcm1 binding site is closer to the start site of the *YJL169W* ORF than to *ASG7*. We did not see any evidence for **a**-specific expression of this gene by RT-PCR (Fig. 1) or Northern blot (data not shown). This result implies that there may be some asymmetry to the mechanism of regulation by this site.

### Repression Mediated by $\alpha 2$ -Mcm1 Target Sites of a Heterologous Promoter

Many of the sites that we identified do not appear to actively regulate gene expression in their endogenous contexts. It is still possible, however, that they are functional binding sites for the  $\alpha 2$ -Mcm1 complex and that they may serve a role in sequestering the  $\alpha 2$  and Mcm1 proteins in the cell. To analyze whether the identified sequences are functional target sites for the  $\alpha 2$ -Mcm1 complex, we measured the level of  $\alpha 2$ -Mcm1-dependent repression mediated by the sites in the context of a heterologous promoter. Oligonucleotides containing the sites were cloned between the UAS and TATA elements in the promoter of a *CYC1-lacZ* reporter. The level of transcriptional repression mediated by each site was determined by comparing the levels of *lacZ* expression from constructs containing the potential  $\alpha 2$ -Mcm1 site to the parent vector without a site in a *MAT $\alpha$*  strain (Table 2). All of the sites in the zero and one-mismatch groups, as well as sites that are most likely to be functional sites in the two-mismatch group, were assayed in the heterologous reporter vector. All of the previously identified sites found in the promoters of the **a**-cell-type-specific *MFA1*, *MFA2*, *STE6*, *STE2*, and *BAR1* genes, as well as sites involved in determining the direction of mating type switching, *DPS1* and *DPS2*, function as repressor sites in the heterologous promoter (Szeto et al. 1997; Zhong and Vershon 1997). The site in the promoter of the **a**-specific *AGA2* gene, which was identified by our search and was previously uncharacterized for binding or repression, also functions as a strong repressor

**Table 1.** Potential  $\alpha 2$ /Mcm1-Binding sites in the Yeast Genome

ORF	Gene <sup>a</sup>	bp	Sequence <sup>b</sup> (TGTANWWWCCNNNNNNGGWWWWNTACA)
<b>mis = 0</b>			
YDR461W	<i>MFA1</i>	204	TGTAA TTACCCAAAAAGGAAATTTACA
YFL026W	<i>STE2</i>	201	TGTAAATTTCCATTTGGGTAA GTACA
YGL032C	<i>AGA2</i>	271	TGTAA TTCCGAATACGGTAA TTACA
YIL015W	<i>BAR1</i>	237	TGTAA TTACCGTAAAGGAAA TTACA
YKL209C	<i>STE6</i>	184	TGTAA TTACCTAATAGGGAAATTTACA
YNL145W	<i>MFA2</i>	223	TGTAT TTACCTATTCGGGAAATTTACA
<b>mis = 1</b>			
YJL170C	<i>DPS1<sup>c</sup></i>	205	TGTAA TTA <u>CT</u> GTATTGGGAAATTTACA
YJL169W	<i>ASG7</i>	166	TGTAAATTTCCCGATGCGGTAA TTAC <u>G</u>
YKL188C	<i>PXA2</i>	259	TGTAA TTA <u>AC</u> GCCGAAGGAAT ATACA
YMR219W	<i>ESC1</i>	29	TGTAA ATTC <u>AC</u> CAAGGGGTTAAGTACA
YMR218C		191	
<b>mis = 2</b>			
YBL108W	<i>DPS2<sup>c</sup></i>	533	TGTAA TTA <u>CT</u> TAAATAGGAA <u>G</u> TTTACA
YEL066W	<i>HPA3</i>	103	TGTAG ATACCGTCTTGGATA GAGCA
YFR007W		55	CGTATATTTCCCAATCCGGTTA TTAAA
YGL075C	<i>MMC1</i>	344	TGAAG TAACCTTGTTAGGTTAATTTCA
YGL187C	<i>COX4</i>	324	TGTCATATACCCGTGAAGGCAT TTACA
YJL062W	<i>LAS21</i>	162	TGTATTTTTCAATAAGAGGATTATAACA
YJL063C	<i>MRPL8</i>	620	TATAGATATGCTTATACGGATATATACA
YJR137C	<i>ECM17</i>	611	TGTACAATACCATCAATGG <u>CTT</u> CTA <u>AA</u>
YJR138W		645	
YLR432W		405	TGTT <u>CATTTCT</u> AACTCTGGTTAATTACA
YLR431C		422	
YNL053W	<i>MSG5</i>	249	TGTAC TTTCCAAAAAGGAAA ATGTA
YNL298W	<i>CLA4</i>	349	GGTAT TTTCCCTTTCCGGAAATATAGA
tS(AGA)	<i>tS(AGA)</i>	469	TGTTTAAATTCCTTGGGGTAT ATACA
YAR020C	<i>PAU7</i>	3'	TGTAG ATACCGTCTACGGATA GAGCA
YPRO09W		3'	TGTAA TTTCCCTCTATGTAAA ATATA
YBR259W		3'	TGTAG TAAATATTATGGAATAGTACA
YIR034C	<i>LYS1</i>	int	TGTAT ATACTTTTAAATGTAAA CTACA
YIR019C	<i>MUC1</i>	int	TGTAA TTTCCACCACTGGTTTTGTACT
YIL131C	<i>FKH1</i>	int	TGTAAATATCCGCTAATGAAATACTACC
YJR091C	<i>JSN1</i>	int	TGGAGATT CCTGTGTTGGAATAGTATA
YAL029C	<i>MYO4</i>	int	TCTAAAATACCAATAAAGGAAA AGACA
YGRWTy3-1	<i>Ty3</i>	int	TGTACAATTCCTGAAAAAGTTT GTATA
YMR146C	<i>TIF34</i>	int	AGTAG TTACCGCATGGGGAAA ATTC <u>A</u>
YML043C	<i>RRN11</i>	int	TGTAC AATGCAAAAGTGGAAAT TT <u>CA</u>
YIL031W	<i>SMT4</i>	int	TGTAT TTTCCAAACCAATCGGTCTACA
YNL250W	<i>RAD50</i>	int	TGAAATTTTCCAAGCGATGAAATTTACA
YGL083W	<i>SCY1</i>	int	TGTT <u>CATA</u> ACCGCGCAAGCTTT GTACA
YNL192W	<i>CHS1</i>	int	TGTAT ATTCCTTTTGT <u>CATTT</u> TT ACA
YNL191W		int	TGTAG AAAC <u>CTT</u> CACCAGGAGA ATACA
YNL041C		int	TGAAT TATCCCTCAGGGCTAA GTACA
YDL074C		int	TGTAGTTTTTACACTTGGATT CTAG <u>A</u>
YDR541C		int	TGTAA AAACCACACTGGGATC GT <u>CCA</u>
YGL186C		int	TGTATTTATCACTTGATGAAAT GTACA
YLR084C	<i>RAX2</i>	int	G <u>GTA</u> ATATTCCATTCTTGGTTT CTATA
YNL047C		int	AGTAC ATTCCTAATGGTAT TG <u>ACA</u>
YNL078W		int	TGGGT TTACCACTTCAGTTATATACA
YDR206W	<i>EBS1</i>	int	TGTACTTTTCCCGAAAAGGAAAT AT <u>CCA</u>
YIR010W		int	AGGAT ATTCCTGAAGAGGATTTTACA
YER037W		int	TGGAT TATCCGTGAAAGGTTTAAATAAA
YPL164C		int	TGTAG TTTTCGAAATGGGATAAATCCA
YJL223C	<i>PAU1</i>	int	TGTAG ATACCGTCTTGGATA GAGCA
YEL049W	<i>PAU2</i>	int	TG <u>CTC</u> TATCCAAGGACGGTAT CTACA
YGL261C		int	TGTAG ATACCGTCTTGGATA GAGCA
YGR294W	<i>MGA1</i>	int	TG <u>CTC</u> TATCCAAGGACGGTAT CTACA
YIL176C		int	TGTAG ATACCGTCTTGGATA GAGCA
YAL068C		int	TGTAG ATACCGTCTTGGATA GAGCA

**Table 1.** (Continued)

ORF	Gene <sup>a</sup>	bp	Sequence <sup>b</sup> (TGTANWWWCCNNNNNNNGGWWWWNTACA)
YBR301W		int	TG <u>CTC</u> TATCCAAGGTCGGTAT CTACA
YHL046C		int	TGTAG ATACCGTCCTTGGATA GAGCA
YKL224C		int	TGTAG ATACCGTCCTTGGATA GAGCA
YDR542W		int	TG <u>CTC</u> TATCCAAGGTCGGTAT CTACA

<sup>a</sup>Blank entries indicate that a gene name has not been assigned to the ORF.

<sup>b</sup>Mismatches with the consensus sequence are underlined. Blank entries indicate a site that is within the promoter region of two divergently transcribed genes. The site is shown on the line above.

<sup>c</sup>The *DPS1* and *DPS2* sites are  $\alpha 2$ -Mcm1 sites located on the left arm of Chromosome III and are involved in mating-type switching but do not appear to regulate a specific ORF (Szeto et al. 1997).

site in the heterologous promoter. The site upstream of *ASG7* also mediates  $\alpha 2$ -Mcm1 repression of the test promoter, although the level of repression, 13-fold, is weaker than many of the sites found in other **a**-specific genes. All other sites tested do not show significant (>3-fold) levels of  $\alpha 2$ -Mcm1-mediated repression and are therefore not functional sites in vivo.

#### Mcm1-mediated Activation of the Potential Regulatory Sites

In the absence of  $\alpha 2$ , the Mcm1 protein binds to  $\alpha 2$ -Mcm1 sites and functions as a transcriptional activator (Bender and Sprague 1987; Jarvis et al. 1988; Ammerer 1990). To test whether any of the sites identified in the search function as Mcm1-dependent activator sites, the *CYC1* UAS elements were removed from the heterologous reporters described above and *lacZ* expression was assayed in a *MAT<sup>a</sup>* strain (Table 2). All of the sites which function as repressor sites in the  $\alpha$  strain also function as activator sites in the *MAT<sup>a</sup>* strain. With the exception of the site in the *MSG5* promoter, most of the sites that do not repress transcription do not function as activator sites. The *MSG5* site, however, exhibits strong Mcm1-mediated activation in both **a** and  $\alpha$  cells, suggesting that this site is a functional Mcm1 site.

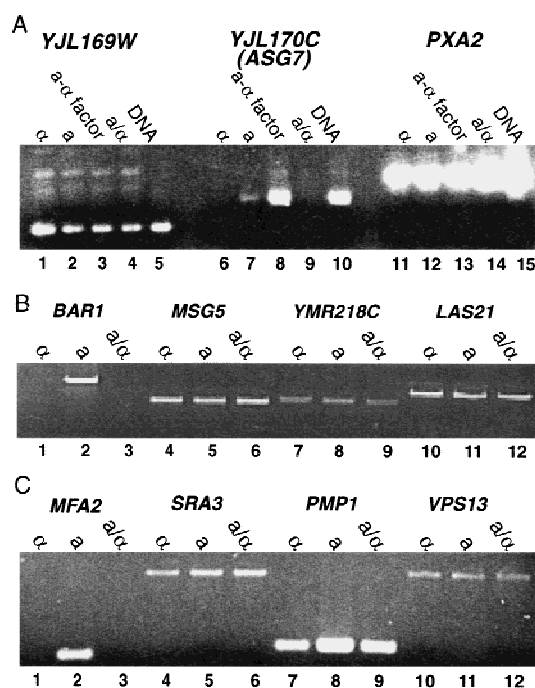
#### In vitro Analysis of the *AGA2* and *ASG7* $\alpha 2$ -Mcm1 Binding Sites

The results described above show that our search has identified two previously uncharacterized, functional  $\alpha 2$ -Mcm1 target sites in the *AGA2* and *ASG7* promoters. To verify that the transcriptional repression and activation mediated by these sites is attributable to binding by the  $\alpha 2$ -Mcm1 complex, the sites were assayed for binding by  $\alpha 2$  and Mcm1 alone, and in combination, by EMSAs (Fig. 3). The *ASG7* site is bound by the  $\alpha 2$  protein although the binding affinity is significantly weaker than that of the *STE6* site. The Mcm1 binding affinity and the cooperative binding by  $\alpha 2$  and Mcm1 to the *ASG7* site, however, are comparable with

the levels observed to the *STE6* site. Under the same conditions, there is no detectable binding by the  $\alpha 2$ -Mcm1 complex to the site found in the *PXA2* promoter. As expected from the in vivo repression data, the  $\alpha 2$  and Mcm1 proteins have similar binding affinity for the site in the *AGA2* promoter as to the *STE6* site (data not shown). The in vitro results are therefore consistent with the in vivo results and show that the *AGA2* and *ASG7* sites function as  $\alpha 2$ -Mcm1 binding and repressor sites.

## DISCUSSION

In this study, we have performed a search of the yeast genome for potential recognition sequences of the  $\alpha 2$ -



**Figure 1** (A) Expression of potential **a**-specific genes in haploid and diploid cells. The expression of *YJL169W* (lanes 1–5), *YJL170C* (lanes 6–10), and *PXA2* (lanes 11–15) were assayed using RT-PCR. RNA prepared from  $\alpha$  (lanes 1, 6, 11), **a** (lanes 2, 7, 12), **a** induced with  $\alpha$  factor (lanes 3, 8, 13), and **a**/ $\alpha$  diploid (lanes 4, 9, 14) strains was used as the template with primers specific for each ORF. Genomic DNA was used as the template to serve as controls for the PCR reactions and the size of the expected fragments (lanes 5, 10, 15). (B) Expression of *BAR1* (lanes 1–3), *MSG5* (lanes 4–6), *YMR218C* (lanes 7–9), and *LAS21* (lanes 10–12) in  $\alpha$  (lanes 1, 4, 7, 10), **a** (lanes 2, 5, 8, 11), and **a**/ $\alpha$  diploid (lanes 3, 6, 9, 12) strains by RT-PCR. (C) expression of genes by RT-PCR that show cell-type-specific expression by microarray assay. *MFA1* (lanes 1–3), *SRA3* (lanes 4–6), *PMP1* (lanes 7–9), and *VPS13* (lanes 10–12) in  $\alpha$  (lanes 1, 4, 7, 10), **a** (lanes 2, 5, 8, 11), and **a**/ $\alpha$  diploid (lanes 3, 6, 9, 12).

**Table 2.** Analysis of the Potential Sites for  $\alpha 2$ -Mcm1-Mediated Repression

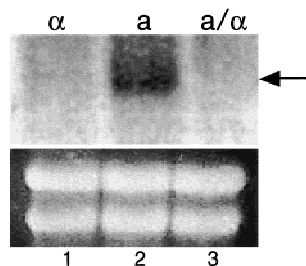
ORF	Gene	bp	Sequence	mRNA <sup>a</sup>	Repression <sup>b</sup>	Activation <sup>c</sup>
<b>mis = 0</b>						
YDR461W	<i>MFA1</i>	204	TGTAA TTACCCAAAAAGGAAATTTACA	<b>a</b>	100×	15×
YFL026W	<i>STE2</i>	201	TGTAATTTCTAATTGGGTAAGTACA	<b>a</b>	33×	28×
YGL032C	<i>AGA2</i>	271	TGTAATTTCCGAATACGGTAATTACA	<b>a</b>	167×	51×
YIL015W	<i>BAR1</i>	237	TGTAATTACCGTAAAAGGAAATTTACA	<b>a</b>	81×	25×
YKL209C	<i>STE6</i>	184	TGTAATTACCTAATAGGGAAATTTACA	<b>a</b>	69×	17×
YNL145W	<i>MFA2</i>	223	TGTATTTACCTATTCGGGAAATTTACA	<b>a</b>	43×	47×
<b>mis = 1</b>						
	<i>DPS1</i>		TGTAATTACGTATTGGGAAATTTACA	<b>a</b>	404×	7.8×
YJL170C	<i>ASG7</i>	205	TGTAATTTCCCGATGCGGTAATTACG	<b>a</b>	13×	14×
YJL169W	ORF	166		<b>a, <math>\alpha</math>, a/<math>\alpha</math></b>		
YKL188C	<i>PXA2</i>	259	TGTAATTAACGCCGAAGGAATATACA	<b>a, <math>\alpha</math>, a/<math>\alpha</math></b>	1.3×	1.5×
YMR219W	<i>ESC1</i>	29	TGTAATTCACCAAGGGGTTAAGTACA	<b>a, <math>\alpha</math>, a/<math>\alpha</math></b>	1.1×	1.2×
YMR218C	ORF	191		<b>a, <math>\alpha</math>, a/<math>\alpha</math></b>		
<b>mis = 2</b>						
	<i>DPS2</i>		TGTAATTACTTAATAGGAAGTTTACA	<b>a</b>	283×	12×
YLR432W	ORF	405	TGTTCATTTCTAACTCTGGTTAATTACA	<b>a, <math>\alpha</math>, a/<math>\alpha</math></b>	2.4×	0.9×
YGL187C	<i>COX4</i>	324	TGTATTTTTCAATAAGAGGATTATAACA	<b>a, <math>\alpha</math>, a/<math>\alpha</math></b>	2.8×	1.0×
YNL053W	<i>MSG5</i>	249	TGTACTTTCCAAAAAGGAAAATGTA	<b>a, <math>\alpha</math>, a/<math>\alpha</math></b>	0.5×	22×
YGL075C	ORF	344	TGTCATATACCCGTGAAGGCATTTACA	<b>a, <math>\alpha</math>, a/<math>\alpha</math></b>	1.3×	0.9×
YJL062W	ORF	162	TATAGATATGCTTATACGGATATATACA	<b>a, <math>\alpha</math>, a/<math>\alpha</math></b>	2.3×	1.0×
YJR138W	ORF	645	TGTACAATACCATCAATGGCTTCTAAA	<b>a, <math>\alpha</math>, a/<math>\alpha</math></b>	2.6×	1.4×

<sup>a</sup>mRNA was examined by Northern blot or RT-PCR in the previous studies and this work. (**a**) The transcript is only present in **a** cells. **a,  $\alpha$ , a/ $\alpha$**  The RNA was present at the same level in haploid **a**,  $\alpha$ , and diploid **a/ $\alpha$**  cells, respectively. The  $\alpha 2$ -Mcm1 sites that are involved in mating-type switching, *DPS1* and *DPS2*, regulate cell-specific expression of several small transcripts (Szeto et al. 1997).

<sup>b</sup>The fold repression is measured assaying the level of expression of a *CYC1-lacZ* reporter plasmid containing the designated  $\alpha 2$ /Mcm1-binding site and comparing it to the parent vector without the site (212 units) in a *MAT $\alpha$*  strain.

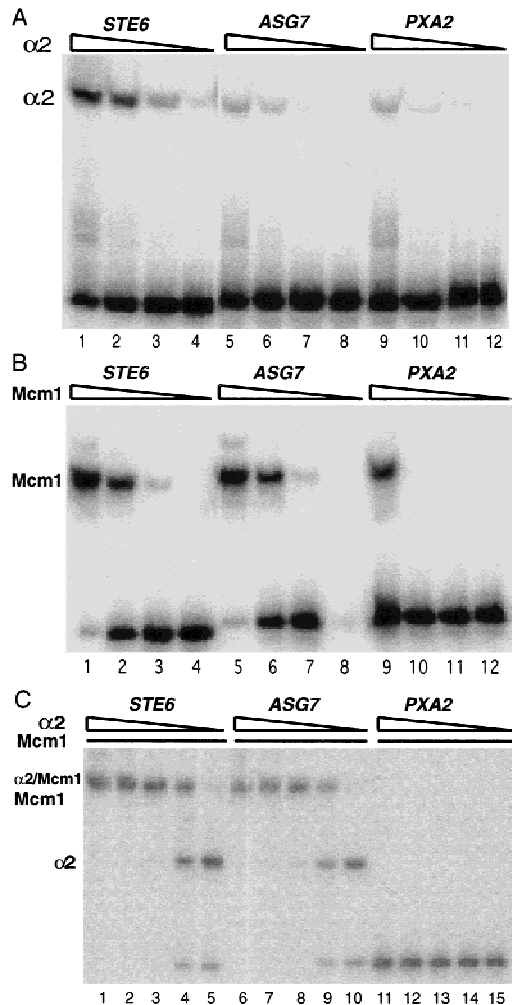
<sup>c</sup>The fold activation is measured assaying the level of expression of a *CYC1-lacZ* reporter plasmid with the designated  $\alpha 2$ /Mcm1-binding site lacking the *CYC1* UAS sites and comparing it to the parent vector without the Mcm1 site (0.4 units) in a *MAT $\alpha$*  strain.

Mcm1 complex. Assuming that the entire yeast genome has an A/T content of 61% (Dujon 1996) and allowing for two mismatches, we expected to find three to four matches in the genome to degenerate TGTAN(A/T)<sub>3-4</sub>CCN<sub>6</sub>GG(A/T)<sub>3-4</sub>NTACA sequence. The fact that we found 59 matches suggests that the presence of these sequences is not random. It is possible that a part of this sequence is used for recognition by other DNA-binding proteins or complexes. For ex-



**Figure 2** Expression of *ASG7* (YJL170C) in haploid and diploid cells. Total RNA was prepared from *MAT $\alpha$*  strain 246.1.1 (lane 1), *MAT $\alpha$*  strain EG123 (lane 2), and the diploid *MAT $\alpha$ /MAT $\alpha$*  strain 246.1.1 X EG123 (lane 3). The blot was hybridized with a radio-labeled DNA fragment specific for the coding region of the YJL170C. An ethidium bromide-stained gel before transfer is shown as a control for RNA loading.

ample, some of the sites could have been identified because they contain an Mcm1-binding site. Mcm1 binds with other cofactors such as the  $\alpha 1$ , Ste12, Arg80, and Sff proteins to regulate other sets of genes (Lydallet et al. 1991; Errede 1993; Hagen et al. 1993; Messenguy and Dubois 1993; Oehlen et al. 1996). One of the sites we identified is in the promoter of *MSG5*, a gene that is highly expressed in response to pheromone (Doiet al. 1994). We have shown that this site is not a functional  $\alpha 2$ -Mcm1 repressor site, but does serve as an Mcm1-dependent activator site and is likely required for the Ste-12-Mcm1 dependent pheromone response. Another explanation for the number of sites we found is that a large number of the sites are within the coding regions of genes or uncharacterized ORFs. Ten of the sites fall within the coding region of the PAU family of proteins, which have an unknown function and contain a low content of serine and threonine residues (Viswanathan et al. 1994). All of the potential  $\alpha 2$ -Mcm1 sites identified in this class of proteins fall in a highly conserved amino acid sequence. These sites were therefore identified in our search because they code for a conserved amino acid sequence and not because they are functional  $\alpha 2$ -Mcm1 binding sites.



**Figure 3** The DNA-binding activity for the ASG7 site in EMSAs. (A) Purified  $\alpha 2$  protein was used in an EMSA with labeled fragments containing the indicated  $\alpha 2$ -Mcm1-binding sites. The protein concentration was  $4.0 \times 10^{-7}$  M (lanes 1,5,9),  $8.0 \times 10^{-8}$  M (lanes 2,6,10),  $1.6 \times 10^{-8}$  M (lanes 3,7,11),  $3.2 \times 10^{-9}$  M (lanes 4,8,12). (B) Purified Mcm1 protein was used in an EMSA with labeled fragments at a concentration of  $1.7 \times 10^{-8}$  M (lanes 1,5,9),  $1.7 \times 10^{-9}$  M (lanes 2,6,10),  $1.7 \times 10^{-10}$  M (lanes 3,7,11), and  $1.7 \times 10^{-11}$  M (lanes 4,8,12). (C) Purified  $\alpha 2$  protein at a concentration of  $8.0 \times 10^{-8}$  M (lanes 1,6,11),  $1.6 \times 10^{-8}$  M (lanes 2,7,12),  $3.2 \times 10^{-9}$  M (lanes 3,8,13),  $6.4 \times 10^{-10}$  M (lanes 4,9,14),  $1.3 \times 10^{-10}$  M (lanes 5,10,15) was mixed with Mcm1<sub>1-96</sub> fragment at a concentration of  $1.7 \times 10^{-9}$  M in the EMSAs. The  $\alpha 2$ -Mcm1 complex includes two molecules of  $\alpha 2$  and two molecules of Mcm1 and the Mcm1 complex includes two molecules of Mcm1. The EMSAs shown are phosphorimages of the gels.

Although the search identified a large number of potential sites in the genome, we have found that relatively few of these sites are functional  $\alpha 2$ -Mcm1 repressor sites. There appears to be a large decrease in the binding affinity and activity between the functional and nonfunctional sites and there do not appear to be any sites with intermediate affinity or activity. Most of the sites identified in the search contain two signifi-

cant mismatches with the search sequence. With the exception of the *DPS2* site, all show substantial decreases in binding affinity in vitro and repression of a heterologous promoter in vivo when compared with the functional sites. In addition, none of the genes near these sites show significant cell-type-specific expression in DNA microarray experiments (Roth et al. 1998). These results further support our findings that these are not functional regulatory sites. It is unlikely that these sites are bound by the protein in vivo. As a result, there do not appear to be any active  $\alpha 2$ -Mcm1 sites in the genome without a specific regulatory target and the presence of an active  $\alpha 2$ -Mcm1 site indicates a functional target gene.

Our search identified two previously uncharacterized  $\alpha 2$ -Mcm1 target sites. One is upstream of *AGA2*, a cell-type-specific subunit of the **a**-agglutinin cell adhesion complex (Cappellaro et al. 1994). Northern results demonstrated that the *AGA2* gene is expressed specifically in **a** cells (de Nobel et al. 1995). Although we expected this gene to be regulated by  $\alpha 2$ -Mcm1, the promoter of this gene had not been analyzed previously. From the search for potential  $\alpha 2$ -Mcm1 sites, we obtained a sequence in the promoter region of *AGA2* that matches the consensus  $\alpha 2$ -Mcm1 site. Furthermore, we have shown that it is a functional  $\alpha 2$ -Mcm1 site in vivo and in vitro. Repression mediated by this site is as strong as  $\alpha 2$ -Mcm1 sites in other **a** cell-type-specific genes (Table 2) and the  $\alpha 2$  and Mcm1 proteins bind to the *AGA2* site with the same affinity as to the *STE6* operator (data not shown). Therefore, as expected, expression of the **a** cell-type-specific *AGA2* gene is directly regulated by the  $\alpha 2$ -Mcm1 repressor complex.

The other novel  $\alpha 2$ -Mcm1-binding site identified in our search is located 205 bp upstream of the translation start site of *YJL170C*. This gene is only expressed in **a** cells and is strongly induced by exposure of the cells to  $\alpha$ -factor (Fig. 1). *YJL170C* is therefore a novel **a** cell-type-specific gene that we have named *ASG7*. A chromosomal null mutation of *ASG7* does not have an effect on the growth rate or mating efficiency (H. Zhong, unpubl.). Preliminary results have shown, however, that *ASG7* is required for inhibition of the **a**-factor mating type receptor, Ste-3 (J. Kim and J. Hirsch, pers. comm.).

Knowledge of the yeast genome sequence makes it possible to examine the expression profile of every gene in the genome under different conditions using DNA microarray techniques (DeRisi et al. 1996; Chu et al. 1998; Roth et al. 1998). This type of analysis has been performed to examine the differential expression of genes in **a** and  $\alpha$  cells (Roth et al. 1998). In agreement with our results and previous findings by others, *MFA1*, *MFA2*, *AGA2*, *STE2*, and *BAR1* show significant differences in expression in **a** versus  $\alpha$  cells. Several

other genes, *PMP1*, *SRA3*, and *VPS13*, which we did not identify in our search, also show higher expression in **a** versus  $\alpha$  cells in the microarray assay. We did not find any potential  $\alpha 2$ -Mcm1 binding sites in the promoter of these genes, however, and did not observe differential cell-type-specific expression of these genes in our assay (Fig. 1C). Two of the genes identified in our search, *STE6* and *ASG7*, did not have large enough differences in the microarray assay to be identified as cell-type-specific genes. Furthermore, the *DPS1* and *DPS2* sites, which have  $\alpha 2$ -Mcm1-dependent roles in the direction of mating type switching, would not be detected in the microarray assay. These sites regulate the cell-type-specific expression of several small transcripts in the left arm of chromosome III that are unlikely to code for any functional protein (Szeto et al. 1997). These transcripts would therefore not likely be detected in microarray assays that monitor expression of well-defined ORFs. Taken together, these results show that each technique identified potential target genes that were not detected by the other assay and that together the two techniques complement. In summary, the search we have performed has identified two previously unidentified sites as well as all of the known  $\alpha 2$ -Mcm1 regulatory sites. This work gives an example that, if the sequence requirement of a DNA-binding protein is relatively well defined, it is possible to search a genome database for its target sites and therefore find novel target genes that are regulated by this protein.

## METHODS

### Northern and RT-PCR Assays

The total RNA from strains 246.1.1 (*Mata*), EG123 (*Mat $\alpha$* ), and 246.1.1 X EG123 (*Mata/ $\alpha$* ) (Silicano and Tatchell 1984) was prepared as described (Ausubel et al. 1997). RNA from **a** cells treated with  $\alpha$ -factor was purified in the same manner except that cells were harvested 2 hr after adding 2 mM of  $\alpha$ -factor at OD<sub>600</sub> of 0.3. RT-PCR assays were performed using the SuperScript Preamplification System kit (GIBCO BRL) according to the manufacturer's instructions. Two micrograms of total RNAs isolated from **a**,  $\alpha$ , or **a/ $\alpha$**  yeast strains were used as templates and oligo(dT)<sub>12–18</sub> was used as the primer to synthesize cDNAs. Pairs of oligonucleotides that specifically anneal to each of the potential genes were used as primers in the second step of PCR for amplification of target cDNAs. These primer pairs were designed to generate ~500-bp fragments. The reactions were run on agarose gels and DNA fragments were visualized with ethidium bromide.

For Northern blot analysis, RNA samples were electrophoresed on a 1% formaldehyde-agarose gel and the RNAs were denatured, neutralized, and blotted to a nitrocellulose membrane. The level of mRNA expression was detected by hybridization with PCR-generated fragments of each gene, which were labeled by random priming with [ $\alpha$ -<sup>32</sup>P]dCTP. Hybridization was performed at 42°C in buffer containing 50% formamide and washed in 2× SSC (Herrick et al. 1990). The membrane was exposed to a phosphor screen and the image was scanned on a Molecular Dynamics phosphorimager.

### $\beta$ -Galactosidase Assays

Transcription reporter plasmids that contain potential  $\alpha 2$ -Mcm1-binding sites were constructed by inserting double-stranded oligonucleotides containing the 31- or 32-bp sites with TCGA overhangs into the *XhoI* site between the UAS and TATA elements in the *CYC1-lacZ* promoter of pTBA23 (Mead et al. 1996). These constructs were used to measure  $\alpha 2$ -Mcm1-dependent repression of the *CYC1* promoter. Plasmids used to measure Mcm1-dependent activation were constructed by digesting the above plasmids with *BglII*, followed by self-ligation (Acton et al. 1997). This step removed the *CYC1* UAS sites from the promoter region, leaving the  $\alpha 2$ -Mcm1 site and *CYC1* and TATA box. The reporter constructs were transformed into the *MAT $\alpha$*  strain 246.1.1 to assay repression or the *MAT $\alpha$*  strain EG123 to assay activation by measuring the level of  $\beta$ -galactosidase activity as described previously (Silicano and Tatchell 1984; Keleher et al. 1988).

### EMSAs

Full-length  $\alpha 2$  and a fragment of Mcm1 (residues 1–98) were expressed in bacteria and purified to >90% homogeneity as described previously (Mead et al. 1996; Acton et al. 1997). The relative  $\alpha 2$ -Mcm1 DNA-binding affinity for the potential  $\alpha 2$ -Mcm1-binding sites were determined by EMSAs as described (Zhong and Vershon 1997).

## ACKNOWLEDGMENTS

We thank J. Gu for help in constructing some of the *LacZ* reporter plasmids and J. Kim and J. Hirsch for communication of unpublished work on the function of *ASG7*. This work was supported by a grant from the National Institutes of Health to A.K.V. (GM49265).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

- Acton, T.B., H. Zhong, and A.K. Vershon. 1997. DNA-binding specificity of Mcm1: Operator mutations that alter DNA-binding and transcriptional activities by a MADS box protein. *Mol. Cell. Biol.* **17**: 1881–1889.
- Ammerer, G. 1990. Identification, purification, and cloning of a polypeptide (PRTF/GRM) that binds to mating-specific promoter elements in yeast. *Genes & Dev.* **4**: 299–312.
- Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidmain, J.A. Smith, and K. Struhl, eds. 1997. *Current protocols in molecular biology*. Wiley/Greene, New York, NY.
- Bender, A.D. and G.F. Sprague. 1987. MAT $\alpha$ 1 protein, a yeast transcription activator, binds synergistically with a second protein to a set of cell-type-specific genes. *Cell* **50**: 681–691.
- Blackwell, T.K. and H. Weintraub. 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A proteins complexes revealed by binding site selection. *Science* **250**: 1104–1110.
- Bult, C.J., O. White, G.J. Olsen, L. Zhou, R.D. Fleischmann, G.G. Sutton, J.A. Blake, L.M. FitzGerald, R.A. Clayton, J.D. Gocayne, A.R. Kerlavage, B.A. Dougherty, J.F. Tomb, M.D. Adams, C.I. Reich, R. Overbeek, E.F. Kirkness, K.G. Weinstock, J.M. Merrick, A. Glodek, J.L. Scott, N.S.M. Geoghagen, and J.C. Venter. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**: 1058–1073.
- Cappellaro, C., C. Baldermann, R. Rachel, and W. Tanner. 1994. Mating type-specific cell-cell recognition of *Saccharomyces cerevisiae*: Cell wall attachment and active sites of **a**- and

- alpha-agglutinin. *EMBO J.* **13**: 4737–4744.
- Chu, S., J. DeRisi, M. Eisen, J. Mulholland, D. Botstein, P.O. Brown, and I. Herskowitz. 1998. The transcriptional program of sporulation in budding yeast. *Science* **282**: 699–705.
- de Nobel, H., J. Pike, P.N. Lipke, and J. Kurjan. 1995. Genetics of  $\alpha$ -agglutinin function in *Saccharomyces cerevisiae*. *Mol. & Gen. Genet.* **247**: 409–415.
- Deckert, G., P.V. Warren, T. Gaasterland, W.G. Young, A.L. Lenox, D.E. Graham, R. Overbeek, M.A. Snead, M. Keller, M. Aujay, R. Huber, R.A. Feldman, J.M. Short, G.J. Olsen, and R.V. Swanson. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**: 353–358.
- DeRisi, J., L. Penland, P.O. Brown, M.L. Bittner, P.S. Meltzer, M. Ray, Y. Chen, Y.A. Su, and J.M. Trent. 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14**: 457–460.
- 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**: 680–686.
- Dickson, D. 1998. Drosophila set for fast-track sequencing. *Nature* **393**: 296.
- Doi, K., A. Gartner, G. Ammerer, B. Errede, H. Shinkawa, K. Sugimoto, and K. Matsumoto. 1994. MSG5, a novel protein phosphatase promotes adaptation to pheromone response in *S. cerevisiae*. *EMBO J.* **13**: 61–70.
- Dujon, B. 1996. The yeast genome project: What did we learn? *Trends Genet.* **12**: 263–270.
- Errede, B. 1993. MCM1 binds to a transcriptional control element in Ty1. *Mol. Cell. Biol.* **13**: 57–62.
- Flanders, D.J., S. Weng, F.X. Petel, and J.M. Cherry. 1998. AtDB, the *Arabidopsis thaliana* database, and graphical-web-display of progress by the Arabidopsis Genome Initiative. *Nucleic Acids Res.* **26**: 80–84.
- Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science* **269**: 496–512.
- 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**: 563–567.
- Hagen, D.C., L. Bruhn, C.A. Westby, and G.F. Sprague Jr. 1993. Transcription of alpha-specific genes in *Saccharomyces cerevisiae*: DNA sequence requirements for activity of the coregulator alpha 1. *Mol. Cell. Biol.* **13**: 6866–6875.
- Herrick, D., R. Parker, and A. Jacobson. 1990. Identification and comparison of stable and unstable mRNAs in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **10**: 2269–2284.
- Herskowitz, I. 1988. Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol. Rev.* **52**: 536–553.
- Jarvis, E.E., D.C. Hagen, and G.F. Sprague Jr. 1988. Identification of a DNA segment that is necessary and sufficient for alpha-specific gene control in *Saccharomyces cerevisiae*: Implications for regulation of alpha-specific and a-specific genes. *Mol. Cell. Biol.* **8**: 309–320.
- Johnson, A. 1992. A combinatorial regulatory circuit in budding yeast. In *Transcriptional regulation* (ed. K.R. Yamamoto and S.L. McKnight), pp. 975–1006. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Johnson, A.D. and I. Herskowitz. 1985. A repressor (MAT $\alpha 2$  product) and operator control expression of a set of cell-type-specific genes in yeast. *Cell* **42**: 237–247.
- Keleher, C.A., C. Goutte, and A.D. Johnson. 1988. The yeast cell-type-specific repressor  $\alpha 2$  acts cooperatively with a non-cell-type-specific protein. *Cell* **53**: 927–936.
- Lashkari, D.A., J.L. DeRisi, J.H. McCusker, A.F. Namath, C. Gentile, S.Y. Hwang, P.O. Brown, and R.W. Davis. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci.* **94**: 13057–13062.
- Lutfiyya, L.L., V.R. Iyer, J. DeRisi, M.J. DeVit, P.O. Brown, and M. Johnston. 1998. Characterization of three related glucose repressors and genes they regulate in *Saccharomyces cerevisiae*. *Genetics* **150**: 1377–1391.
- Lydall, D., G. Ammerer, and K. Nasmath. 1991. A new role for Mcm1 in yeast: Cell cycle regulation of *SWIS* transcription. *Genes & Dev.* **5**: 2405–2419.
- Mead, J., H. Zhong, T.B. Acton, and A.K. Vershon. 1996. The yeast  $\alpha 2$  and Mcm1 proteins interact through a region similar to a motif found in homeodomain proteins of higher eukaryotes. *Mol. Cell. Biol.* **16**: 2135–2143.
- Messenguy, F. and E. Dubois. 1993. Genetic evidence for a role for MCM1 in the regulation of arginine metabolism in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**: 2586–2592.
- Miller, A.M., V.L. MacKay, and K.A. Nasmyth. 1985. Identification and comparison of two sequence elements that confer cell-type specific transcription in yeast. *Nature* **314**: 598–603.
- Oehlen, L.J., J.D. McKinney, and F.R. Cross. 1996. Ste12 and Mcm1 regulate cell cycle-dependent transcription of *FAR1*. *Mol. Cell. Biol.* **16**: 2830–2837.
- Passmore, S., G.T. Maine, R. Elble, C. Christ, and B.-K. Tye. 1988. *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MAT $\alpha$  cells. *J. Mol. Biol.* **204**: 593–606.
- Roth, F.P., J.D. Hughes, P.W. Estep, and G.M. Church. 1998. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.* **16**: 939–945.
- Silicano, P.G. and K. Tatchell. 1984. Transcription and regulatory signals at the mating-type locus in yeast. *Cell* **37**: 969–978.
- Smith, D.L. and A.D. Johnson. 1992. A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an alpha 2 dimer. *Cell* **68**: 133–142.
- . 1994. Operator-constitutive mutations in a DNA sequence recognized by a yeast homeodomain. *EMBO J.* **13**: 2378–2387.
- Strathern, J., J. Hicks, and I. Herskowitz. 1981. Control of cell type in yeast by the mating type locus. The  $\alpha 1$ - $\alpha 2$  hypothesis. *J. Mol. Biol.* **147**: 357–372.
- Szeto, L., M.K. Fafalios, H. Zhong, A.K. Vershon, and J.R. Broach. 1997. Alpha2p controls donor preference during mating type interconversion in yeast by inactivating a recombinational enhancer of chromosome III. *Genes & Dev.* **11**: 1899–1911.
- Tan, S. and T.J. Richmond. 1998. Crystal structure of the yeast MAT $\alpha$ 2/MCM1/DNA ternary complex. *Nature* **391**: 660–666.
- Tatchell, K., K.A. Nasmyth, B.D. Hall, C. Astell, and M. Smith. 1981. In vitro mutation analysis of the mating-type locus in yeast. *Cell* **27**: 25–35.
- Treisman, R. 1995. DNA-binding proteins. Inside the MADS box. *Nature* **376**: 468–469.
- Treisman, J., P. Gonczy, M. Vashishtha, E. Harris, and C. Desplan. 1989. A single amino acid can determine the DNA-binding specificity of homeodomain proteins. *Cell* **59**: 553–562.
- Venter, J.C., M.D. Adams, G.G. Sutton, A.R. Kerlavage, H.O. Smith, and M. Hunkapiller. 1998. Shotgun sequencing of the human genome. *Science* **280**: 1540–1542.
- Viswanathan, M., G. Muthukumar, Y.S. Cong, and J. Lenard. 1994. Seripauperins of *Saccharomyces cerevisiae*: A new multigene family encoding serine-poor relatives of serine-rich proteins. *Gene* **148**: 149–153.
- West, A.G., P. Shore, and A.D. Sharrocks. 1997. DNA binding by MADS-box transcription factors: a molecular mechanism for differential DNA bending. *Mol. Cell. Biol.* **17**: 2876–2887.
- Wolberger, C., A.K. Vershon, B. Liu, A.D. Johnson, and C.O. Pabo. 1991. Crystal structure of a MAT $\alpha 2$  homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* **67**: 517–528.
- Zhong, H. and A.K. Vershon. 1997. The yeast homeodomain protein Mata2 shows extended DNA-binding specificity in complex with Mcm1. *J. Biol. Chem.* **272**: 8402–8409.

Received March 24, 1999; accepted in revised form September 3, 1999.