CHAPTER THREE

MASTER REGULATORY TRANSCRIPTION FACTORS CONTROL CELL TYPE

In Chapter 1, I introduced the two haploid cell types of yeast, mating type **a** and mating type α. The third cell type is produced by mating of **a** and α cells, the diploid **a**/α type. We learned in Chapter 2 of two important differences between the **a** and α cell types: Each produces a distinct mating pheromone and each displays a receptor on its cell surface that specifically detects the pheromone produced by the opposite mating type. We also learned that the **a**/α cell type is incapable of mating, but is able to undergo meiosis and sporulation.

The Pattern of Gene Expression Distinguishes the **a**, α, and **a**/α Cell Types

These differences in behavior among the **a**, α, and **a**/α cell types are caused by different patterns of gene expression. For example, **a** cells—but not α cells—transcribe the genes encoding the **a**-specific pheromone (**a**-factor), and the receptor for the α-specific pheromone (the α-factor receptor Ste2). Likewise, α cells—but not **a** cells—transcribe the genes encoding α-pheromone and the **a**-pheromone receptor, Ste3. In **a**/α cells, which do not mate, neither the pheromones nor the receptors are produced because the genes that encode them are not transcribed.

Gene Regulation

Because the differences between **a**, α, and **a**/α cells involve the control of gene transcription, I will mention briefly some general features of gene regulation in eukaryotes that will be illustrated in this chapter.

Genes are controlled by regulators that bind to specific DNA sequences; in general, these sequences are found near the genes they regulate. A given regulator can either turn on (activate) or turn off (repress) the transcription of a gene to which its binds. As we will see in this chapter, regulators are often composed of several different polypeptide subunits.

The archetypal regulator has two domains: a DNA-binding domain that recognizes a specific sequence and a domain that is required for the regulator to control transcription. Thus, regulators that activate transcription ("activators") each contain a "transcriptional activation domain," whereas those that repress transcription ("repressors") each contain a "transcriptional repression domain." The function of activation and repression domains is to bring proteins to the promoter that more generally control transcription. Activation domains, for example, recruit protein complexes that in turn recruit RNA polymerase II, the enzyme responsible for synthesizing messenger RNA (mRNA) in eukaryotes.

With these principles in mind, we will spend the remainder of the chapter focusing in detail the mechanisms by which cell type–specific transcription of genes occurs.

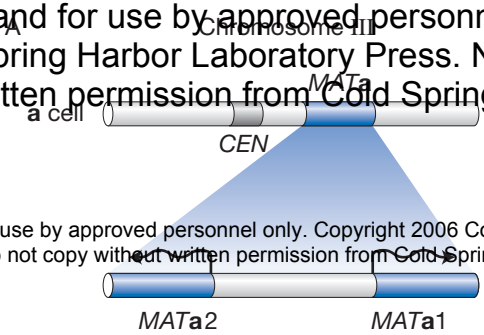
The *MAT* Locus Is the Master Controller of Cell Type and Occurs in Two Versions

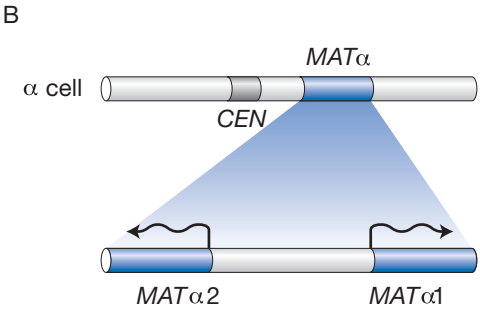
The distinct patterns of gene expression between **a** and α cells are determined by a single genetic locus, called *MAT* (Fig. 3-1). The *MAT* locus differs between the two cell types (Fig. 3-1) and is the only difference between the genomes of the two mating types. **a** cells have the *MAT***a** version of the mating-type locus, whereas α cells have the *MAT*α version.

Genes at the *MAT* locus encode proteins that bind specific DNA sequences. Two proteins, called **a**1 and **a**2, are encoded by the *MAT***a** locus (Fig. 3-1A). These are different from the two proteins encoded by the *MAT*α locus, which are named α1 and α2 (Fig. 3-1B). In **a**/α cells, both versions of *MAT* exist (Fig. 3-1C).

The Overall Scheme for Mating-type Regulation

The proteins encoded by the *MAT* locus associate with other proteins to form the regulators whose actions lead to the patterns of transcription that are characteristic of each cell type. Figure 3-2 provides a "cheat sheet" that summarizes the overall scheme of gene regulation in **a**, α, and **a**/α cell types. It may be useful for the reader to refer back to Figure 3-2 to assist in placing the individual pieces of information described below into the overall scheme.





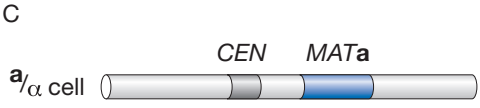


Figure 3-1. The mating-type locus, *MAT*.

α1 Is Required for the Activation of α-specific Genes

In this section, we will focus on genes that are expressed solely in α cells-—α-specific genes (α-sgs). α1 binds directly to the promoters of α-sgs by recognizing a specific DNA sequence present in the promoters of this class of genes. In the absence of α1, α-sg transcription is not activated and the genes lie dormant. Since **a** cells do not contain the gene encoding α1, they cannot express α-sgs.

α-sgs include two redundant genes that encode α-factor, *MF*α1 and *MF*α2 (not to be confused with *MAT*α1 and *MAT*α2!), as well as the **a**-factor receptor gene, *STE3* (Fig. 3-3).

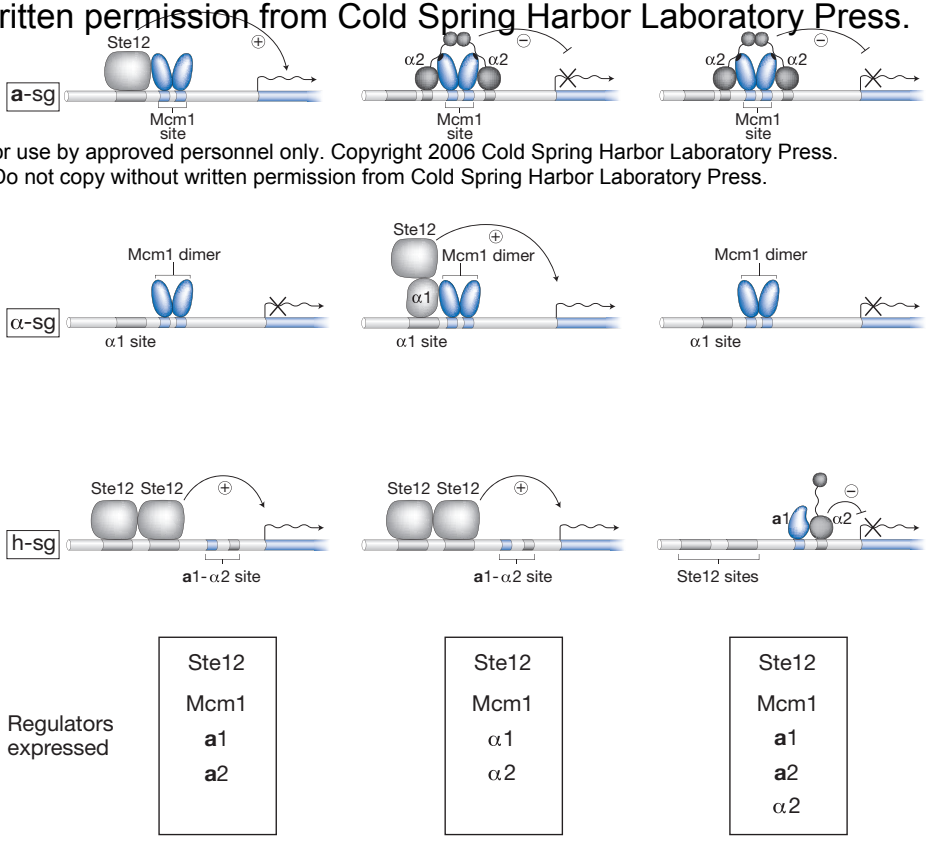


Figure 3-2. Overall scheme of cell type control. Shown is how **a**-specific genes, α-specific genes, and haploid-specific genes are regulated in **a**, α, and **a**/α cells. Also shown are the cell types in which the regulators are expressed. Note that one particular class of haploid-specific genes is shown—those that respond to the extracellular presence of mating pheromone. Also note that of the proteins shown, only Ste12 contains a domain capable of activating transcription.

Although the regulatory proteins encoded by the *MAT* locus are ultimately responsible for specifying cell type, they do not work alone. As shown in Figure 3-4, α1 binds to DNA along with two other proteins called Mcm1 and Ste12. These two proteins are expressed in all three cell types. The Mcm1 protein binds cooperatively as a dimer to a DNA site adjacent to the α1 binding site (Fig. 3-4). Cooperative binding is explained in Box 3-1. α1 and Mcm1 exhibit DNA binding cooperativity with each other as well.

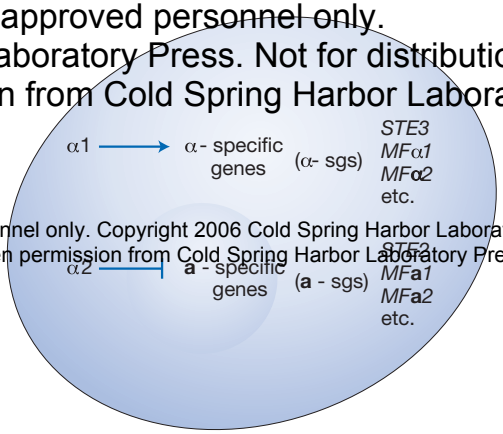


Figure 3-3. Control of cell type–specific genes by α1 and α2.

Binding to DNA and activation of transcription can be mediated by different polypeptides. For example, neither α1 nor Mcm1 are capable of activating transcription on their own—a third protein is required for α-sgs to be transcribed. This protein is Ste12, and among the three proteins specifically necessary for the expression of α-sgs, it is the only one that possesses a transcriptional activation domain. Ste12 does not contact the DNA directly at the promoters of α-sgs, but is recruited to the promoter by a protein–protein interaction with α1 (Fig. 3-4). Once recruited through this interaction, Ste12 activates transcription of α-sgs.

To summarize, α1 activates the transcription of α-sgs such as *STE3*. It does so by binding to DNA cooperatively with the Mcm1 protein and recruiting a third protein called Ste12.

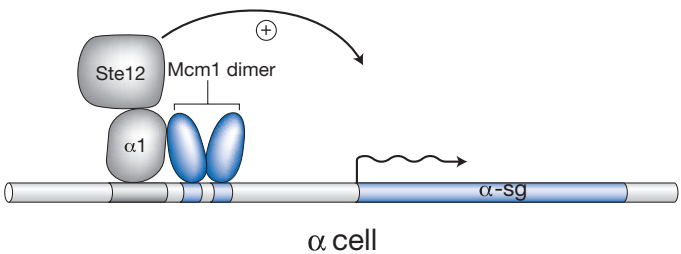


Figure 3-4. Activation of α-sgs by α1.

Box 3-1. Cooperative Binding of Proteins to DNA: Definition, Mechanism, and Consequences

Cooperative binding of proteins to DNA is a common occurrence. By definition, it occurs when the presence of one DNA-binding protein increases the concentration required by another protein to bind to DNA. The proteins can be two molecules of the same protein (as in the case of the cooperative binding of Mcm1 subunits to DNA) or different proteins (in the case of α1 binding cooperatively to DNA with Mcm1).

In its simplest manifestation, cooperative DNA binding between two proteins typically depends on the following features: (1) that the two proteins bind to sites that are linked spatially on the DNA, (2) that the two proteins can interact with each other, and (3) that the concentrations of the proteins, and their affinities for DNA, are low enough that their binding to each other becomes necessary for the DNA to be occupied by one or both proteins.

What are the consequences of cooperative DNA binding?

One of them has been mentioned earlier in the chapter: Cooperativity allows for combinatorial control. What do I mean by this? By making the binding to DNA of one regulator depend, through cooperativity, on the binding of another, a given gene can be set to "switched on" (for example) only when both regulators are present. If each regulator is available to bind DNA only in response to a specific signal, then the gene is switched on only when both signals are present. This can be extended to more signals by making the binding of further regulators also depend on cooperativity.

By mixing and matching the DNA-binding sites for different regulators (which are able to bind cooperatively) within promoters of different genes, new combinations of signals can be required to switch on different genes, allowing a promoter to integrate signals.

Cooperative DNA binding can also be used to generate steep "all-or-none" effects. That is, the binding of a protein to DNA can be exquisitely sensitive to its concentration, with small changes in that concentration having dramatic effects on DNA site occupancy. Thus the state of gene expression, stable in one state, can be poised to completely switch to an alternative stable state over a very narrow change in regulator concentration. Although this property is not relevant to the discussion of the mating-type regulators described in this chapter, it is crucial for understanding gene regulation in other contexts, such as the genetic switch of phage λ.

Cooperativity also helps deal with an issue that arises from the fact that DNA-binding proteins not only bind to specific DNA sequences, but to other "nonspecific" DNA sequences, albeit with a lower affinity. This presents a problem for a given protein trying to find its site because the number of nonspecific sites in a genome is typically huge compared to the number of specific sites for that protein. Thus, even though the affinity of each nonspecific site is low—and thus each site holds the protein for only a very short time—the overall effect of the population of nonspecific sites can be immense. In effect, the protein may spend the vast majority of its time caught up in an endless sampling of the low-affinity sites.

Cooperativity overcomes this problem. Because of the large number of nonspecific sites, and because the protein samples each so fleetingly, it is unlikely that two molecules of the protein will simultaneously occupy adjacent nonspecific sites. Specific sites, with their higher affinity for the protein, hold that protein for longer, thus vastly increasing the chance that protein bound at one such site will make contact with another molecule of protein bound at an adjacent specific site (if such is available). The two proteins can then bind there cooperatively, stabilizing each other at those sites.

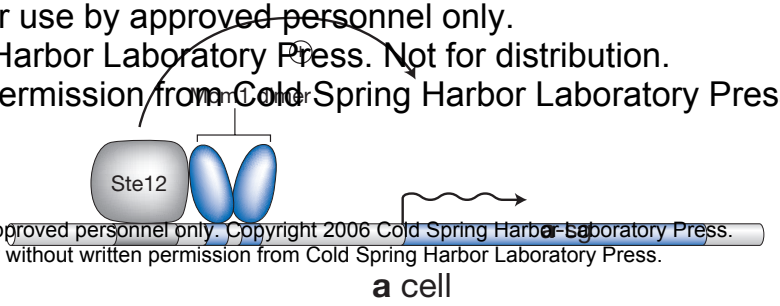


Figure 3-5. Activation of **a**-sgs by Ste12-Mcm1.

In **a** Cells, **a**-specific Genes Are Activated by a Ste12-Mcm1 Complex

In *MAT***a** cells, neither **a**1 nor **a**2 plays a role in the activation of **a**-specific genes. Rather, the activator protein Ste12 binds directly to DNA sites that exist in the promoters of **a**-sgs. These sites each occur adjacent to a binding site for the Mcm1 dimer, facilitating the cooperative binding of Ste12 and Mcm1 to the promoters of **a**-sgs (Fig. 3-5). This occurs in much the same way as α1 and Mcm1 bind to α-sg promoters in α cells, with the key difference being that **a**-sgs contain a DNA sequence recognized by Ste12 rather than the sequence recognized by α1 (Fig. 3-5).

From what we have learned so far, we can see that Ste12 can interact specifically with a number of different molecules. These include a specific DNA sequence present in the promoters of **a**-sgs, the proteins Mcm1 and α1, and molecules involved in the activation of transcription. The ability of Ste12 to bind to these different molecules is mediated by distinct segments of the protein.

α2 Is Part of a Repressor of **a**-specific Genes

If Ste12 and Mcm1 are expressed in both **a** and α cells, what prevents **a**-sgs from being expressed in α cells? The answer is that the α2 protein is a repressor that turns off **a**-sgs in α cells by binding to their promoters (Figs. 3-3 and 3-6). Like α1 and

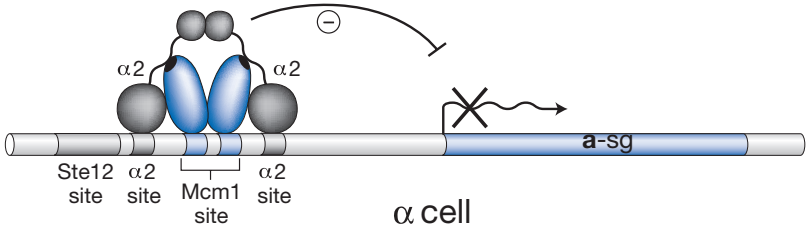


Figure 3-6. Repression of **a**-sgs by α2.

Ste12, α2 binds directly to DNA cooperatively with Mcm1, but it recognizes a different DNA sequence (Fig. 3-6). Specifically, two molecules of α2 bind, one to either side of the Mcm1 dimer. They not only contact Mcm1 but also contact each other (Fig. 3-6). Thus, Mcm1 functions at both **a**-sgs and α-sgs, binding to different partners in the different cell types. Its primary function appears to be to provide a cooperative DNA-binding partner for α1, Ste12, or α2.

α2 Represses Transcription by Recruiting a General Corepressor

So how does α2 turn off **a**-sgs? A simple mechanism that one could imagine is that α2 occludes the site on Mcm1 that interacts with Ste12, thereby preventing Ste12 from binding to DNA cooperatively with Mcm1 (as implied in Fig. 3-6). Although this mechanism may play a role at some **a**-sgs, it does not fully explain how α2 acts to repress transcription. To accomplish repression, α2 brings additional proteins to the promoter. Specifically, it recruits a "corepressor complex" comprised of the proteins Ssn6 and Tup1 (Fig. 3-7). This complex causes the repression of transcription when brought to promoters that would otherwise be active. The Ssn6-Tup1 complex blocks the recruitment of RNA polymerase II to **a**-sgs in α cells, but the exact mechanisms by which it does so are not understood.

Tup1 is also involved in the repression of many other genes in yeast. For each class of genes, Tup1 is brought to DNA by a distinct DNA-binding repressor protein analogous to α2 (Fig. 3-8). Proteins related to Tup1 and involved in gene repression are found in multicellular organisms. For example, in the fruit fly, *Drosophila melanogaster,* a gene called *groucho* encodes a Tup1-like protein involved in regulating gene expression during development. As in yeast, it is recruited by a variety of DNA-binding repressor proteins to repress transcription. Flies lacking *groucho* die during embryogenesis and exhibit abnormal development of the nervous system.

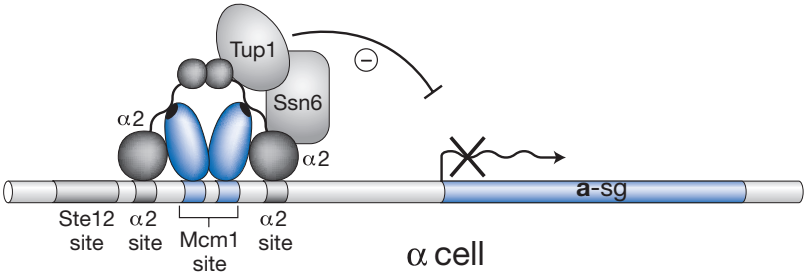
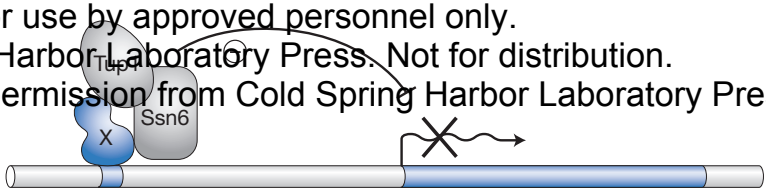


Figure 3-7. α2 represses genes by recruiting the Tup1-Ssn6 corepressor.



X	target genes
Crt1	DNA damage-induced genes
Mig1	glucose-repressed genes
Rox1	oxidative stress genes

Figure 3-8. The Tup-Ssn6 corepressor is recruited to distinct gene sets by different DNA-binding proteins symbolized by "x." Each DNA-binding protein is inactivated by a particular environmental cue. For example, Crt1 normally represses a set of genes encoding proteins involved in the cellular response to DNA damage, and DNA damage causes the inactivation of Crt1, resulting in the transcription of these genes.

In **a**/α Cells, **a**1 (Encoded by *MAT***a**) and α2 Form a Repressor of Haploid-specific Genes

In this section, we will focus on the mechanisms that specify the diploid **a**/α cell type. As with the haploid **a** and α cell types, it is the pattern of gene expression that gives the **a**/α cell type its specialized characteristics.

As discussed in Chapter 2, **a**/α cells lack the ability to mate, but have the ability to undergo meiosis and sporulation. Because they are produced by mating between *MAT***a** and *MAT*α cells, **a**/α cells could in principle express all four genes encoded by the two mating-type loci. However, this is not the case. Instead, **a**/α cells express only three of the genes: α2, **a**1, and **a**2. Because α1 is not expressed, α-sgs remain dormant, and because α2 is expressed, **a**-sgs are also repressed.

How is α1 turned off? This is where the **a**1 protein comes in. Although it is expressed in **a** cells, **a**1 has no function in these cells. However, in **a**/α cells, **a**1 binds cooperatively with α2 to specific DNA sites that consist of a sequence recognized by **a**1 adjacent to a sequence recognized by α2. This **a**1-α2 complex functions as a repressor. It binds to a site in the promoter of the α1 gene, shutting off its transcription in **a**/α cells (Fig. 3-9A).

The **a**1-α2 repressor also binds to the promoters of another class of genes, the haploid-specific genes (h-sgs) (Fig. 3-9B). The h-sgs are defined as genes that are expressed in **a** cells and α cells, but not in **a**/α cells. Many h-sgs are involved in

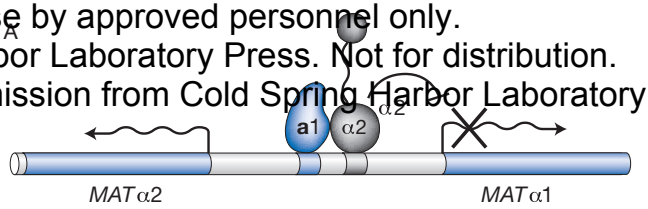


Figure 3-9. Repression of haploid-specific genes by **a**1-α2. (*A*) The **a**1-α2 complex represses the transcription of α1. (*B*) The **a**1-α2 complex represses the transcription of haploid-specific genes.

the mating responses of both **a** and α cells. These genes are activated by Ste12 bound to their promoters (see Fig. 3-2); their regulation will be covered further in Chapter 5. Because **a**/α cells express neither the pheromone nor the pheromone receptor genes (because of the presence of α2 and the absence of α1), they cannot mate. And even if these genes were to be expressed, mating would be blocked because of the repression of h-sgs by the **a**1-α2 repressor.

A Haploid-specific Regulator Explains Why **a** and α Cells Do Not Undergo Meiosis and Sporulation

A key h-sg is *RME1*, which stands for *R*epressor of *Me*iosis *1* (Fig. 3-10). As its name suggests, expression of *RME1* in **a** and α cells prevents the expression of genes that are required for meiosis and sporulation. Because *RME1* is repressed by

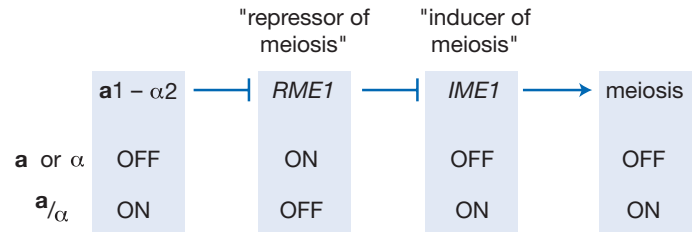


Figure 3-10. **a**1-α2 allows **a**/α cells to undergo meiosis by repressing the transcription of *RME1*, a repressor of meiosis. Shown are the expression states of *RME1* and its target gene *IME1* in **a** or α vs. **a**/α cells. Note that *IME1* expression requires that cells be starved for nitrogen- and carbon-containing nutrients.