A regulatory hierarchy for cell specialization in yeast

Ira Herskowitz

The specialized sets of genes that determine different cell types in yeast are controlled by combinations of DNA-binding proteins some of which are present only in certain cell types whereas others are present in all cell types. Final differentiation requires an inductive signal that triggers both gene transcription and cell-cycle arrest. Synthesis of the proteins coded by the 'master regulatory' mating-type locus is regulated so as to generate a heterogeneous mitotic cell population containing a stem-cell lineage.

ALTHOUGH budding yeast is a unicellular organism, it exhibits many of the processes that are crucial to the development of multicellular organisms, such as cell specialization and cell–cell interactions. The relative simplicity of yeast and its experimental tractability have made it possible to synthesize an integrated view of how cell specialization in yeast results. This view encompasses transcriptional regulatory proteins as well as response to intercellular growth and differentiation factors.

In a multicellular organism, the different types of specialized cell play different roles in the intact organism. In yeast, the specialized cells play different roles during the life cycle of the organism[1-3]. There are three different types of yeast cell: the **a** and α mating types are typically haploid and are specialized for mating with each other. Mating entails the fusion of the two haploid cells to produce the third cell type, the **a**/α diploid cell, which can undergo meiosis and sporulation when nutritionally starved.

There are many differences among these cell types. Perhaps most notably, the **a** and α cells each have a distinctive cell-signalling system that enables mating to take place between the two partners. The α cells produce an extracellular peptide, α-factor, which acts on a receptor produced only by **a** cells; **a** cells likewise produce an extracellular peptide, **a**-factor, that acts on a receptor produced only by α cells. **a**/α cells are incompetent for mating and neither produce nor respond to either factor.

All of the differences between these cell types are determined by the two alleles of the mating-type locus, *MAT*: cells with the *MAT***a** allele exhibit the **a** mating type, cells with the *MAT*α allele exhibit the α mating type, and cells with both *MAT* alleles exhibit the **a**/α phenotype. These alleles turn out to encode components of regulatory proteins that occupy a key position in yeast cell specialization: they govern expression of specialized gene sets and sit at the top of the regulatory hierarchy that generates a specific mitotic cell lineage.

I begin with an explanation of what we know about how the products of the **a** and α alleles determine yeast cell types. We shall see that these products do not act alone and shall identify their molecular co-participants and their regulators. Furthermore, we shall see that cell specialization results from a combination of cell-autonomous programming (by these regulatory proteins) and environmental response (induction), two recurring themes in classical developmental biology. The molecular information obtained from studies of yeast may be directly applicable to understanding the parts played by these processes in development of multicellular eukaryotes and how different combinations of transcriptional regulators can give rise to distinctive cell types.

Differential gene expression

Four classes of genes are defined on the basis of their distinct patterns of expression in the three yeast cell types (Table 1). α-specific genes are expressed only in α cells; **a**-specific genes are expressed only in **a** cells; haploid-specific genes are expressed in both haploid cell types but not in diploid (**a**/α) cells; and sporulation-specific genes are expressed only in **a**/α diploid cells (and then only after nutritional starvation). These differential patterns of expression are directly determined by the products of the *MAT* locus (except for sporulation-specific genes; discussed below).

The *MAT* locus codes for three polypeptides[4,5] that are components of regulatory proteins: **a**1 is encoded by the *MAT***a** allele and is thus produced only in **a** cells; α1 and α2 are encoded by the *MAT*α allele and are produced only by α cells. These products are responsible for three regulatory activities[4]: α1, α2 and **a**1 − α2. Figure 1 shows how these three activities programme expression of the four classes of genes. α1 is a positive regulator of transcription of the α-specific genes—it is needed for their expression. α2 is a negative regulator of transcription of the **a**-specific genes—it blocks transcription of these genes. Thus, in an α cell, the α-specific genes are expressed whereas ***a***-specific genes are not; hence, the cell displays an α phenotype. The **a**1 protein produced in an **a** cell does not play any role in programming the **a** cell phenotype. Rather, the appropriate gene set (**a**-specific genes) is expressed because of the absence of the negative regulator α2, and the inappropriate gene set (α-specific genes) is not expressed because of the absence of the positive regulator α1. **a**1 does have an important role in **a**/α cells, where it associates with α2 (ref. 6; C. Goutte and A. Johnson, personal communication) to comprise a novel negative regulatory protein, **a**1 − α2, that is responsible for many

TABLE 1 Cell-type-specific gene sets

Gene set	Representative members	Cell type **a**	α	**a**/α
a-specific genes	*STE2, STE6, BAR1*	+	−	−
α-specific genes	*STE3*	−	+	−
Haploid-specific genes	*STE12, HO, RME1*	+	+	−
Sporulation-specific genes	*IME1, SPO13*	−	−	+

a-specific genes include those coding for the structural genes for **a**-factor (*MFA1* and *MFA2*) and a protein necessary for its secretion (*STE6*)[105], the receptor to α-factor (*STE2*), and a protease that degrades α-factor (*BAR1*). α-specific genes include those coding for the structural genes for α-factor (*MF*α*1* and *MF*α*2*) and the receptor to **a**-factor (*STE3*). Haploid-specific genes include those necessary for mating (*STE4, STE5, STE12* and *STE18*), nuclear fusion (*KAR1*), mating-type interconversion (*HO*), regulation of meiosis (*RME1*), and also for transcription of the retrotransposon Ty1 (refs 36, 106, 107). The genes coding for subunits of the G protein—*SCG1/GPA1, STE4* and *STE18* (ref. 108)—are all haploid-specific genes. Sporulation-specific genes are expressed only after nutritional starvation and include an activator of early events in meiosis (*IME1*; ref. 21), genes necessary for proper meiotic chromosome segregation (*SPO12* and *SPO13*), and the sporulation-specific glucoamylase[109]. Additional references can be found in refs 1 and 2.

of the properties of this cell type. In particular, **a**1-α2 turns off synthesis of α1 (refs 7, 8), which eliminates transcription of the α-specific genes. **a**1-α2 also turns off expression of the large group of haploid-specific genes, which are otherwise expressed in both haploid cell types.

Our understanding of control of yeast cell type has come from a wealth of *in vivo* and *in vitro* observations. First of all, mutants defective in the different products of *MAT* exhibit defects in regulation of the target gene sets: for example, α cells that lack α1 fail to express α-specific genes[9]; mutants that lack α2 show the inappropriate behaviour of expressing **a**-specific genes[10]; mutants defective in either **a**1 or α2 are unable to turn off haploid-specific genes[11,12]. Specific DNA sequences located in the upstream regions of the different gene sets have been identified that confer cell-type-specific expression on reporter genes[13-16]. Finally, the polypeptides encoded by *MAT* have been shown to be components of site-specific DNA-binding proteins which bind to upstream regulatory regions of their target genes[6,15,17-19]. These studies show that the proteins that bind to the specialized genes *in vitro* are responsible for regulation of these genes *in vivo*.

The sporulation-specific genes differ from the other classes in that they are not directly regulated by the products of the *MAT* locus. **a**1-α2 does not bind directly to the sporulation-specific genes, but activates them by turning off synthesis of an inhibitor of their expression[12,20]. This inhibitor is the product of the *RME1* gene, which blocks an early step in meiosis[21] (see Fig. 1).

Mechanisms of activation and repression

α1 and α2. The products of the *MAT* locus do not act on their own to produce the cell-type-specific pattern of gene expression in haploid cells, but interact with a common protein (originally termed PRTF[18] or GRM[17]) that is present in all cell types. The alphabet-soup of different names for the same factor has been simplified (or made more complex, depending on one's outlook) by the recent finding that PRTF/GRM is encoded by the *MCM1* gene[22-24]. I shall refer to it hereafter as MCM1. This protein has binding sites in the upstream regulatory regions of both the **a**-specific and the α-specific genes[17-19] and seems to be a transcriptional activator of both gene sets[13,17,25,26]. It is the α1 and α2 products of *MAT* that determine whether transcriptional activation takes place (Fig. 2). Research on how they do this has revealed molecular mechanisms that are likely to underlie other examples of eukaryotic cell specialization.

In **a** cells, MCM1 is able to bind to its site in the upstream region of **a**-specific genes and is thus able to stimulate transcription. MCM1 cannot bind to its targets in α-specific genes to activate these genes (for reasons explained below). Thus, in **a** cells, **a**-specific genes are expressed and α-specific genes are not.

In α cells, MCM1 must be helped to activate the α-specific genes and prevented from activating the **a**-specific genes. These are functions of α1 and α2 proteins, respectively (see Fig. 2). α1 might assist MCM1 in two ways. First of all, α1 (which binds to its target site only in the presence of MCM1) enhances binding of MCM1 to its binding site (the 'P' site)[13,18,19,23]. Although this might be the sole role of α1, studies *in vitro* of binding of α1 and MCM1 to P sites have led Tan *et al.*[19] to raise an additional possibility. They suggest that MCM1 has different conformations depending on the site to which it binds. When it binds to the P site in **a**-specific genes, it is in the active conformation for stimulating transcription. By contrast, binding of MCM1 to a P site in an α-specific gene does not induce the active conformation; this results only when α1 binds adjacent to it. Whatever α1 is doing, an elegant and informative analysis by Sprague and colleagues[13,18] shows that it is MCM1, and not α1, that is the activator. A synthetic P site ('P-PAL') was constructed that is a perfect palindromic version of naturally occurring P sites from different α-specific genes. This site allows binding of MCM1 *in vitro* without α1 and, when used as an upstream activating sequence for a reporter gene, it allows expression in all cell types—in other words, in the absence of α1. These observations indicate the MCM1 is responsible for activating α-specific genes and that α1 merely assists MCM1.

As noted by Jarvis *et al.*[24], control of α-specific genes by MCM1 and α1 provides a way to think about how cell-type-specific regulation might evolve from a constitutively expressed gene. In the first step, the binding site for a general transcriptional activator (analogous to MCM1) becomes enfeebled by mutation. The reduced binding ability is subsequently compensated for by a second protein (analogous to α1), which is present in some cell types but not in others. It is also possible to imagine that activator proteins such as the oestrogen receptor could adopt helper proteins to assist them in binding to weak, imperfect-palindromic sites[27].

How does MCM1 stimulate transcription? MCM1 contains an acidic region (ref. 26; see also refs 23, 24) and thus may stimulate transcription in the same way as proposed for other activator proteins containing an 'acid blob', by direct contact with a component of the transcription machinery[28]. The nucleotide sequence of *MCM1* also shows that it is similar to a protein that is thought to be a transcriptional regulator in mammalian cells, the serum response factor (SRF), which binds to the upstream region of the *fos* gene[29]; moreover, MCM1 is able to bind to the SRF binding site *in vitro*[23,30], suggesting strong conservation of DNA-binding motifs in phylogeny. The functional consequences of the binding of MCM1 to its natural sites

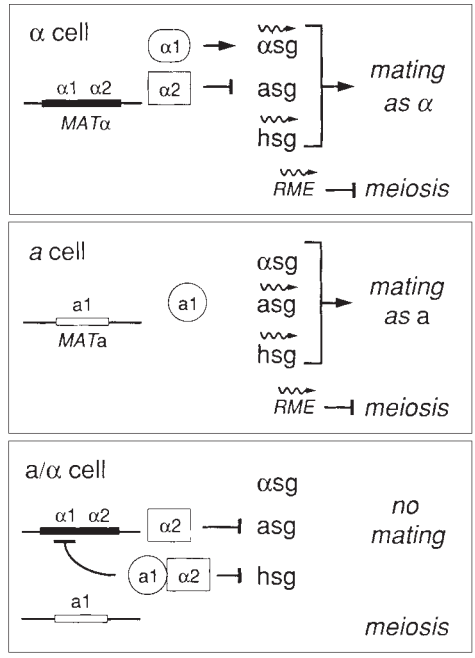

FIG. 1 Expression of cell-type-specific gene sets is governed by regulatory proteins coded by the mating-type locus. The three panels show how the regulatory proteins coded by the mating-type locus (α1, α2 and **a**1-α2) govern transcription of three different gene sets (*αsg*, α-specific genes; *asg*, **a**-specific genes; and *hsg*, haploid-specific genes). Transcription is indicated by a wavy arrow. α1 activates transcription of α-specific genes; α2 inhibits transcription of **a**-specific genes; **a**1-α2 inhibits transcription of haploid-specific genes. *RME1* is a haploid-specific gene whose product inhibits initiation of meiosis (probably by repression of the *IME1* gene; ref. 21) and thus expression of the *ssg* (sporulation-specific) gene set. As a result of action of these regulatory proteins, the three cell types exhibit the indicated phenotypes with respect to mating and meiosis. Members of the different gene sets are listed in Table 1. Nucleotide sequences of *MATα* and *MAT***a** are given in ref. 5. *MATα* contains 747 bp of DNA (drawn as a black rectangle) that is not present in *MAT***a**; *MAT***a** contains 642 bp of DNA (drawn as an open rectangle) that is not present in *MATα*.

in yeast cells, however, are profoundly influenced by both the DNA and the cellular context of the binding, in a way that is likely to be true of many DNA-binding proteins in specialized eukaryotic cells.

Thus, in the upstream regions of **a**-specific genes, there are $\alpha 2$ binding sites on both sides of the binding site for MCM1; $\alpha 2$ binds to these sites to prevent MCM1 from activating the transcription of **a**-specific genes. There are two ways in which $\alpha 2$ might repress transcription[17]. $\alpha 2$ may 'mask' MCM1 (for example, by covering up its activation domain) so that it cannot make contact with other parts of the transcription machinery. Another possibility is that the MCM1-$\alpha 2$ complex is able to contact the transcription machinery, but 'locks' it in a nonproductive mode. The masking explanation is similar to that invoked for the yeast negative regulator, GAL80, which binds to the activator GAL4, and is thought to cover up the GAL4 activation domain[31]. Although the masking model for $\alpha 2$ is appealing, it does not explain how MCM1-$\alpha 2$ can repress transcription when it acts upstream of an intact transcriptional activation region[15].

$\alpha 2$ (and **a1**) contain a segment that is similar to the homoeodomain found originally in *Drosophila* proteins[32,33] and that has since been discovered in other families of DNA-binding proteins[34]. Keleher *et al.*[17] suggest that the properties of $\alpha 2$ may provide an answer to how these proteins, which have relatively low specificity of binding, recognize their target sites. Just as MCM1 increases the affinity of $\alpha 2$ for its operator at least 50-fold[17], an MCM1 analogue might enhance the ability of homoeodomain proteins to recognize their target sites. Perhaps the conserved region adjacent to the homoeodomain of the POU family of proteins[34] contacts such a companion protein.

a1–α2. $\alpha 2$ is versatile: in association with MCM1, it represses transcription of **a**-specific genes; but in conjunction with **a1** it represses the haploid-specific genes. This regulatory species, **a1**-$\alpha 2$, is present only in **a**/α cells and is a molecular signature for this cell type. In the presence of **a1** polypeptide, $\alpha 2$ exhibits a new binding specificity[6], recognizing the **a1**-$\alpha 2$ operator sequence[16,35], which does not contain an MCM1 binding site (C. Goutte and A. Johnson, personal communication; and ref. 6). Although one might have imagined that all $\alpha 2$ subunits in an **a**/α cell would be present as the heteromeric **a1**-$\alpha 2$ species, this is not the case[6]: **a**/α cells contain both species of repressor, $\alpha 2$ and **a1**-$\alpha 2$, which recognize distinctive operator sites to turn off two different gene sets (the **a**- and haploid-specific genes).

It is likely that **a1**-$\alpha 2$ works by preventing other proteins, such as transcriptional activators, from binding to DNA. Genes repressed by **a1**-$\alpha 2$ comprise a diverse collection that uses different activator proteins. For example, **a1**-$\alpha 2$ blocks transcription of the Ty1 element[36] and the *HO* gene[11,16], which require different transcriptional activators[37,38].

Combinatorial control. Molecular programming of the specialized cell types of yeast demonstrates the way in which combinations of polypeptide subunits generate different regulatory activities. Table 2*a* shows the distribution of subunits in the three different cell types. In addition to listing **a1**, $\alpha 1$, $\alpha 2$ and MCM1, I include the STE12 protein, which is discussed below. These subunits interact with each other in a pairwise way to generate several distinctive regulatory species (Table 2*b*): $\alpha 1$-MCM1, $\alpha 2$-MCM1, **a1**-$\alpha 2$ and MCM1-STE12. A comparison of the subunits (Table 2*a*) with the regulatory activities resulting from their pairwise association (Table 2*b*) shows clearly how combinations of different subunits (such as members of the POU family) can generate cell-specific gene expression[34].

Induction of differentiation

Although **a** and α cells in pure culture exhibit many of their characteristic cell specializations, such as production of mating factors and receptors, some specializations are seen only when **a** and α cells are cultured together. A particularly striking example of this is expression of the *FUS1* gene, the product of which is necessary for cell fusion during mating. Its synthesis is induced at least 500-fold (refs 39, 40). In the presence of cells of opposite mating type, synthesis of the mating factors and receptors—in fact, all of the **a**-specific and α-specific genes that have been tested—is also induced, in this case roughly 3- to 5-fold (refs 25, 41, 42; S. Michaelis, K. Kubo and I.H., unpublished observations; summarized in ref. 2). It is the mating factors that are responsible for this induction. α-factor, a peptide of 13 amino acids, is secreted from cells and is freely diffusible. **a**-factor is a peptide of 12 amino acids whose last residue is a farnesylated cysteine that also contains a methyl ester[43]. Although **a**-factor is found extracellularly[44], it is conceivable that **a**-factor is anchored to the **a** cell membrane by the farnesyl group and is presented to α cells by cell–cell contact rather than by diffusion.

The mating factors have several additional effects on responding cells (see ref. 2), all of which serve to coordinate and facilitate mating. This induction can be viewed as the final differentiation of a yeast cell, which occurs only when the mating partner is present. Another effect of the mating factors is that they induce responding cells to arrest cell division in G1 (refs 44, 45). Thus **a**-factor and α-factor are negative growth factors. They may work by interfering with the activity of the CDC28 protein

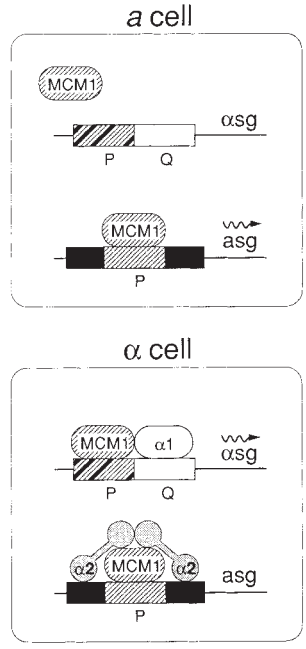

FIG. 2 Transcriptional control of α- and **a**-specific genes by $\alpha 1$, $\alpha 2$ and MCM1 (aka PRTF, GRM). Appropriate expression of α-specific and **a**-specific genes in an **a** cell (upper panel) and in an α cell (lower panel) is mediated by interactions of $\alpha 1$ and $\alpha 2$ with MCM1 protein in the upstream regulatory regions of the target genes. The MCM1 binding site is labelled 'P' and the $\alpha 1$ binding site 'Q'. The P sites are drawn differently for α- and **a**-specific genes to indicate that MCM1 does not bind in the same way to both. In the **a** cell, MCM1 binds to its site in the upstream region of the **a**-specific genes and stimulates their transcription. It is unable to bind effectively to the P site in the upstream region of α-specific genes. In the α cell, $\alpha 2$ binds next to MCM1 in the upstream region of *asg* and thereby represses their transcription. $\alpha 1$ helps MCM1 to activate transcription of α-specific genes by assisting its binding or by inducing it to adopt a conformation suitable for activating transcription. $\alpha 2$ binds as a dimer to the sites drawn in black on opposite faces of the double helix[17,114]. Each $\alpha 2$ monomer is composed of two domains[114]: the C-terminal domains bind to DNA, and the N-terminal domains contact each other and MCM1 protein. MCM1 is inferred to be a dimer. The diagram is not drawn to scale: an $\alpha 2$ monomer is 215 amino acids and an MCM1 monomer is 285 amino acids, based on nucleotide sequence information for these genes[5,26]. Adapted from refs 13 and 18 according to information from refs 17 and 114.

kinase[46,47] and thereby cause arrest in G1 at 'start'[48], as occurs for mutants defective in the *CDC28* gene. Yeast cells thus exhibit a choice of either proliferation or differentiation depending on the presence of the mating factors. Cells can be induced by mating factors at stages other than G1 (ref. 41). Thus the link between cell-cycle arrest and differentiation may be due simply to the induction by the mating factors of a protein required to arrest the cell cycle (a candidate is the product of the *FAR1* gene; F. Chang and I. H., unpublished data).

Induction of gene expression by mating factors provides an opportunity to trace how a signal passes from the cell surface to the nucleus, where it stimulates gene expression. Although there are gaps in our understanding of the pathway, many of the genes and proteins involved have been identified (reviewed in refs 2, 49), and the outlines of a pathway can be seen (Fig. 3). The mating-factor receptors themselves, products of the *STE2* and *STE3* genes, are members of a family of receptors that includes rhodopsin and the beta-adrenergic receptor, all of which communicate with a tripartite G protein with alpha, beta and gamma subunits. Binding of the peptide ligands, **a**-factor or α-factor, to their receptors apparently causes the beta and gamma subunits of the G protein to dissociate from the alpha subunit and then trigger subsequent events through an unknown intracellular signal[50]. Induction of gene expression by the mating factors can occur in the presence of cycloheximide, indicating that synthesis of new proteins is not required for induction[40,41,51]. Protein kinases are likely to be involved in the pathway of signal transduction: the *STE7* and *STE11* gene products, which are required for response, show sequence similarity to protein kinases (ref. 52; B. Errede, personal communication) and act downstream of the G protein, although their precise roles and their substrates are not known. The trail becomes indistinct at this point but picks up later, in the nucleus.

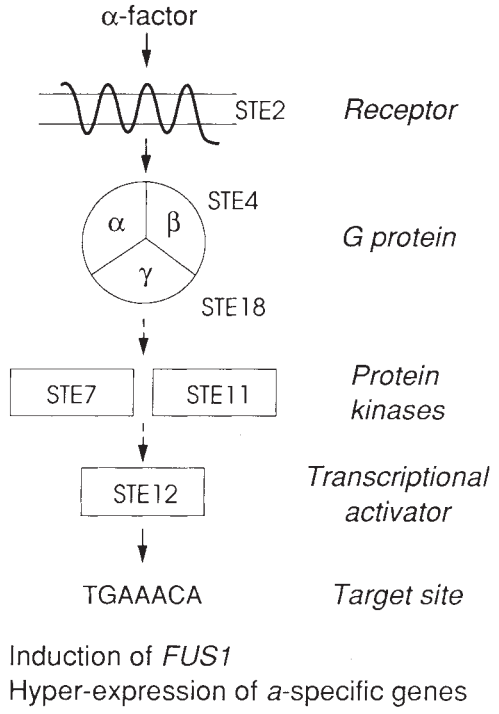

FIG. 3 Induction of differentiation by mating factors: the pathway of signal transduction from membrane to nucleus.

The diagram shows the pathway of signal transduction for **a** cells stimulated by α-factor. As described in the text, an activated receptor (an integral membrane protein located in the plasma membrane) causes the subunits of the G protein to dissociate. This ultimately leads to activation of STE12 protein, which (either by itself or with help from the MCM1 protein; see text) binds to the 'induction box' sequence in the upstream regulatory region of various genes, such as *FUS1* and the members of the **a**-specific gene set, to trigger transcription of these genes. Arrest of the cell cycle may result from transcriptional activation of another of these genes by STE12 (F. Chang and I. H., unpublished data; ref. 54). The manner in which STE12 becomes activated is not known, but it is speculated that the dissociated G protein in some way (indicated by a dotted arrow) activates the STE7 and STE11 proteins (which are believed to be protein kinases), which ultimately lead to activation of the STE12 protein. The alpha, beta and gamma subunits are encoded by genes *GPA1/SCG1* (refs 50, 110, 111) *STE4*, and *STE18* (ref. 108), respectively. Inactivation of any of the *STE* genes shown in the figure disrupts the response pathway—mutants neither arrest nor exhibit gene induction. Inactivation of the *SCG1/GPA1* gene leads to constitutive expression of the pathway, that is, gene induction and arrest even in the absence of α-factor. The pathway for signal transduction for α cells is the same as shown, except that the receptor is coded by the *STE3* gene; all of the intracellular machinery is identical in **a** and α cells.

A DNA sequence that confers inducibility by the mating factors and a protein that apparently binds to this site have been identified. A short sequence (TGAAACA—the 'induction box' or PRE, 'pheromone response element') is necessary for inducible response (ref. 25; see also refs 39, 51). These sites are present in different numbers of copies in the upstream regions of inducible genes. For **a**- and α-specific genes, these sites are located in different positions with respect to other regulatory sites—notably binding sites for MCM1 protein. The STE12 protein has recently been shown to bind to this sequence[53] and presumably is ultimately responsible for induction. An appealing hypothesis[53,54] is that activity of the STE12 protein is governed by phosphorylation under control of the upstream protein kinases (Fig. 3). Other yeast activator proteins (heat-shock factor[55] and GAL4[56]) are known to be phosphorylated under different conditions.

STE12 plays a part not only in induced expression but also in setting the basal level of expression of the **a**- and α-specific genes[57,58]. Mutants defective in *STE12* exhibit modest but significant reductions (3- to 10-fold) in expression of **a**- and α-specific genes for strains grown in pure culture. Thus STE12 is needed for full basal expression but is not an absolute requirement. (The *STE7* and *STE11* genes are also required for basal expression[40,42,58], presumably by contributing to the basal activity of STE12.) A recent finding helps put the role of STE12 in perspective. Errede and Ammerer[54] have found that STE12 and MCM1 proteins interact: binding of STE12 to the upstream region of an **a**-specific gene requires MCM1. (This is the rationale for including MCM1-STE12 as another regulatory activity in Table 2.) It is not known whether STE12 can activate transcription by itself or whether it always requires interaction with another protein such as MCM1 (ref. 54). In any event, there seem to be three contributors to the expression of **a**- and α-specific gene sets: the most important is MCM1 protein (which, for α-specific genes, also requires α1); it is probably required under all conditions. The second contributor is STE12 protein in its 'unstimulated' state; it may stabilize MCM1 or otherwise enhance transcription and thereby contribute to basal expression. The third contributor is STE12 protein in the 'stimulated' state that results from activation of the signalling pathway. It is possible that the induced level of expression of **a**-specific and α-specific genes results from enhancement of the basal mode of expression.

We can now see that cells specialized for mating achieve their final, fully differentiated state by a combination of two processes: intrinsic programming and response to the environment. The final differentiation of an **a** or α cell is thus triggered at the appropriate time—when a mating partner is in the vicinity. In the context of developmental biology, it might be appropriate to view cells that have been induced by mating factors as distinct cell types (H. R. Horvitz, personal communication). An additional cell type is generated by induction(nutritional starvation) of the **a**/α cell. By this method of reckoning, there are six cell types of yeast (Table 3).

Triggering differentiation at the appropriate time is also important during development of multicellular organisms. The

TABLE 2 Distribution of polypeptide subunits and regulatory activities in three different yeast cell types

a Polypeptide subunits

Cell type	**a**1	α1	α2	MCM1	STE12
a	+	−	−	+	+
α	−	+	+	+	+
a/α	+	−	+	+	−

b Regulatory activities

a	α	**a**/α
MCM1	MCM1	MCM1
	MCM1-α1	
	MCM1-α2	MCM1α2
MCM1-STE12	MCM1-STE12	**a**1-α2

The indicated gene products are present in different cell types as shown in *a*. They interact with each other to form the regulatory activities shown in *b*. It is likely that STE12 can also act independently of MCM1, as dimers of STE12 binding sites are sufficient to create a mating-factor-inducible upstream activation site (S. Fields, personal communication).

embryological concept of induction (see ref. 59) is increasingly becoming accessible to molecular study at the level of the single cell. Differentiation of photoreceptor cells in the *Drosophila* ommatidium is thought to be a series of such inductive events, the final one being induction of the R7 precursor cell by the adjacent R8 cell[60,61]. Another example comes from the nematode, in which the anchor cell induces differentiation of nearby vulval precursor cells[62].

The process by which yeast cells mate demonstrates the phenomena of refinement and reinforcement, in which interacting cells generate a commitment to each other. As we have noted, cells express a basal level of mating pheromone and pheromone receptor. Thus a 'naive' α cell triggers a nearby **a** cell to synthesize more α-factor receptor (ref. 42), thereby increasing its responsiveness to α-factor, and also to synthesize more **a**-factor[63]. This **a**-factor then stimulates the α cell to synthesize more **a**-factor receptor[41] and more α-factor and so on. Obviously this process of mutual reinforcement (positive feedback) must be limited in some way. What happens, of course, is that the end product of the process—the **a**/α cell formed by mating—turns off the machinery for synthesis of the mating pheromones and receptors. Similar signalling events may occur during *Drosophila* development (reviewed in ref. 64). For example, at borders of stripe patterns seen in the early embryo, adjacent cells might mutually induce each other to produce a unique differentiated state.

Regulatory hierarchy

As described above, the products of the *MAT* locus are regulatory proteins that determine cell type. Similarly, the products of the homoeotic genes of *Drosophila* determine the particular cell types present in a body segment (ref. 65; reviewed in ref. 64). One of the important questions in unravelling the molecular basis of development is to understand what determines which regulatory proteins are present or active in a given cell. In yeast, the *MAT* gene products are part of a regulatory hierarchy (Fig. 4). The *MAT* locus is at the 'top', coding for proteins that trigger expression of the specialized gene sets; and it is in turn controlled by the product of the *HO* gene, which determines which set of regulatory proteins is expressed in a given cell. The *HO* gene is itself regulated in several different ways.

Regulation of *MAT* by *HO*. The product of the *HO* gene does not encode a direct transcriptional regulator but operates a 'cassette' mechanism determining which *MAT* allele shall occupy the mating-type locus[66,67] (Fig. 5). Although they are haploid, **a** and α cells in fact contain the information for both *MAT* alleles, with only one allele actually occupying the *MAT* locus in any one cell. Only the allele at the *MAT* locus is active: two additional *MAT* genes, one copy of *MAT***a** and one of *MAT*α, are silent. The *HO* gene encodes a site-specific endonuclease[68] that initates a genetic rearrangement in which the active allele at the mating-type locus is replaced by a copy of the opposite mating-type allele stored at the silent loci. As a result, the information from the silent copy is activated, and the mating type of the cell is switched. Thus, although *HO* encodes an endonuclease, its position in the regulatory hierarchy is formally that of a regulator of the master regulatory locus.

Regulation of *HO*. The hierarchy continues. The *HO* gene itself is subject to three different kinds of control: it is expressed only in **a** and α cells[11] (cell-type control); it is expressed only in late G1 (ref. 69) (cell-cycle control); and it is expressed only in certain cells of a mitotic lineage, mothers but not daughters[70] (mother–daughter or asymmetric control). *HO* is expressed only when these three conditions are met. The latter two types of control are responsible for generating mitotic cell lineages in which cells change from α to **a** and the reverse according to specific rules[71] (Fig. 6*a*). One of the most striking observations is that the population is heterogeneous with regard to ability to switch mating types: some cells (mother cells) are competent to switch, whereas other cells (daughter cells) are not. Yeast cell divisions therefore generate a population containing a stem cell lineage, in which the original cell type is preserved and novel cell types are generated. Further analogies are shown in Fig. 6*b*.

The products of thirteen different genes are responsible for carrying out these different types of regulation. They act ultimately in an upstream region of the *HO* gene comprising 1,500 base pairs (bp), which has been subdivided into two broad regions known as upstream regulatory sequences (URSs)[72]: URS1 is responsible for mother–daughter control, and URS2 is responsible for cell-cycle control (Fig. 7). Regulation is brought about by **a**1-α2 (refs 11, 16) and the products of six *SWI* ('switch')[38,73] and five *SIN* ('SWI-independent') genes[74,75]. The *SWI* gene products are all necessary for transcription of *HO*, and they are therefore formally positive regulators of *HO*. The *SIN* products are thought to be negative regulators of *HO* expression. Some of these may be antagonized by SWI products.

The mechanism responsible for cell-type negative regulation of *HO* is straightforward: the *HO* upstream regulatory region is sprinkled with recognition sites for **a**1-α2: ten sites are present throughout URS1 and URS2 (refs 16, 72). Binding of **a**1-α2 to these sites presumably prevents activation events described below from occurring.

The sequences responsible for cell-cyle control of the *HO* gene have been identified. The URS2 region contains ten copies of a 'cell-cycle box', the conserved motif of which is CAC-

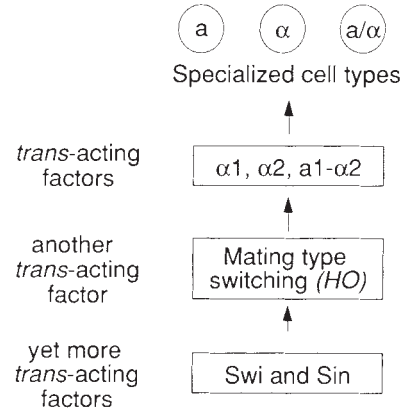

FIG. 4 Regulatory hierarchy for specialized cell types in yeast.

GAAAA (ref. 69). A DNA-binding activity that recognizes the cell-cycle box has also been identified (CCBF, for 'cell-cycle box factor')[76]. SWI4 protein is present in the complex[115]; SWI6 is required for CCBF activity and may also be part of the complex. The basis for limiting *HO* expression to late G1, after the 'start' event of the cell cycle, is not yet known, but it may be that CCBF is phosphorylated by the CDC28 protein kinase, which regulates 'start'[46,48]. Another possibility is that CDC28 phosphorylates another part of the transcription machinery, such as a subunit of RNA polymerase[77].

The restriction of *HO* expression to mother cells appears to be due to the behaviour of the regulatory proteins SWI5 and SIN3. The SWI5 product, which is required for transcription of *HO*[38], binds to a site in the URS1 region through its three zinc fingers[78,79]. As SWI5 is not required in the absence of SIN3 (ref. 74), it is inferred that SWI5 activates transcription by antagonizing SIN3, which otherwise inhibits *HO* expression in some way. The important role of SWI5 and SIN3 in mother–daughter control is deduced from the observations that daughter cells can express *HO* if they are defective in *SIN3* (refs 74, 75) or if they overexpress the *SWI5* gene[80]. It thus seems that mother and daughter cells differ from each other in their level of SIN3 activity, which is high in daughter cells and low in mother cells. The current working hypothesis is that this difference results from mother cells having a higher level of SWI5 than daughter cells, but the reason for this difference is not yet known. Preferential localization of SWI5 to mother cells or proteolytic degradation of SWI5 in daughter cells have been invoked as possible explanations.

Combinatorial control of *HO* expression. *HO* expression occurs in **a** or α cells only when two conditions are met: that the cells are mothers and that they are in the G1 phase of the cell cycle. Our working hypothesis (B. Andrews, W. Kruger, C. Peterson and I.H., manuscript in preparation) for this combinatorial control of *HO* involves sequential activation of the two upstream regulatory domains by two independent events, one occurring only in mother cells, the other only in late G1 (Fig. 7). In this scheme, SWI5 (through its action in URS1) initiates a sequence of events that culminates in the binding of CCBF to the cell-cycle box sequences in URS2. First, SWI5 'sweeps the deck', clearing the URS1 region of inhibitory factors such as SIN3. This allows another protein, which we call SWI1, 2, 3 (because it requires the *SWI1*, *SWI2* and *SWI3* genes) to act in URS1. Because SIN1 protein prevents the activation of transcription from the cell-cycle boxes located in URS2 (B. Andrews and W. Kruger, unpublished data), we have proposed that SWI1, 2, 3 activates transcription by antagonizing SIN1. Thus, in the next step in the sequence, SWI1, 2, 3 overcomes repression by SIN1. Then, CCBF is able to rush onto the scene, bind to the cell-cycle boxes and trigger transcription of *HO*. Given such an exhausting and complex scenario, it is no wonder that this process can occur only once each cell division and then only in mother cells. The particular form of complex regulation of the *HO* gene can thus be explained simply by control of the access of an activator protein (CCBF in this case) to its DNA binding site.

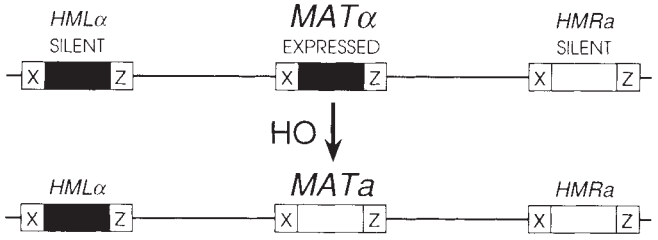

FIG. 5 Changing alleles of the mating-type locus by genetic rearrangement. The top line shows the arrangement of cassettes on chromosome III in an α cell. The cassette at *MAT* is expressed; those located at *HML* and *HMR* are repressed by Sir (see ref. 112). Switching to **a** occurs by removing the α cassette from *MAT* and replacing it by information from the silent version of *MAT***a**, which is located at *HMR***a** (refs 66, 67). This process is initiated by the product of the *HO* gene, which is an endonuclease that produces a double-strand break at *MAT*[68,113]. Subsequent repair of this break using the homologous sequences located at X and Z leads to a duplicative transposition of information from *HML* or *HMR* to *MAT*. The central regions of the cassettes (shown as solid or open rectangles) are nonhomologous segments described in the legend to Fig.1; the X and Z regions are ~700 and ~200 bp respectively. More details on the structure of the cassettes are given in ref. 5.

TABLE 3 Six cell types of yeast

Cell type	Properties
α cell	has the potential to mate
α cell induced by **a**-factor	has the ability to mate
a cell	has the potential to mate
a cell induced by α-factor	has the ability to mate
a/α cell	has the potential to sporulate
a/α cell induced by nutritional signal (starvation)	sporulates

Although the molecular analysis of the *HO* upstream region is far from complete, a few additional points are worth making. First of all, the ability of the cell-cycle box element to function as a transcriptional activation site is strongly dependent on its location. When removed from URS2 and assayed in an artificial construct, it functions as an activating sequence[73]. But in its normal context as part of the URS2 region, it functions as an activating sequence only if the URS1 region is present: if URS1 is deleted, activation from the cell-cycle box is eliminated[72]. This difference results from inhibition of the cell-cycle box in URS2 by SIN1, which, as we have seen, is itself regulated by events in URS1. I expect that there will be many such examples of regulatory elements from eukaryotic genes that exhibit this position-dependent behaviour. The second point that I would like to make is a speculation on how the complex *HO* regulatory region may have evolved. The present-day upstream region may have evolved by the juxtaposition of a URS1-like region to a DNA segment containing multiple cell-cycle boxes. This would yield a regulatory region that has tandem transcriptional activating segments, analogous to the regulatory region of the yeast *CYC1* gene[81]. In the case of *HO*, the two segments of the upstream region have become functionally linked, with the SIN1 protein acting as the coupling device. It operates in both regions, limiting access to the cell-cycle boxes in URS2, but itself being controlled by events that take place in URS1.

Several of the proteins that are involved in regulating *HO* are related to proteins found in other eukaryotes. Thus, the CDC28 protein kinase is related to the cdc2 protein kinase of fission yeast[82], which is itself similar to a component of the general inducer of M phase, maturation promoting factor[83]. Breeden and Nasmyth[84] identified a widely distributed motif coding for 33 amino-acid residues in *SWI6* that is present in the *cdc10* gene of fission yeast, as well as in the *Drosophila Notch* and nematode *lin-12* genes. The SWI6-cdc10 motif, whose function is unknown, has now been found in the *SWI4* gene[115] and in the nematode *glp-1* gene[85].

Concluding comments

The role of negative regulation in cell-type-specific gene expression. Cell-type-specific expression of yeast genes involves an interplay of positive and negative regulatory factors. The MCM1 protein is likely to be an important transcription factor for yeast, with many sites throughout the genome. As we have seen, it is responsible for transcription of two specialized gene sets, the α- and **a**-specific genes, and it may also be necessary for transcription of essential genes such as *CDC28* (refs 23, 26). The positive factor that distinguishes α cells from **a** cells is the α1 protein, which assists MCM1 in activating transcription of

the α-specific gene set. The negative factor that distinguishes α and **a** cells is α2. MCM1 protein is fully capable of binding to the upstream regions of **a**-specific genes, even in α cells. But binding of MCM1 does not lead to transcription of these genes because α2 flanks it and thereby prevents MCM1 from functioning. Repression of transcription by α2 thus goes hand-in-hand with transcriptional activation by α1 in programming cell-type-specific gene expression.

Yeast cell specialization involves a second repressor activity, **a**1-α2. Although this negative regulator also contains the α2 polypeptide, it binds to a different site and almost certainly represses by a mechanism different from α2 (preventing proteins from binding rather than inhibiting their ability to activate transcription). Negative regulators such as α2 and **a**1-α2 provide an explicit demonstration of why the presence of a transcriptional activator protein may not be sufficient to activate gene expression. We think that there will be numerous other examples of negative factors acting at the level of DNA. The similarity between MCM1 and SRF makes it likely that analogues of α2 (and α1) will also be found in mammalian cells. Genetic evidence for negative regulators in *Drosophila* has recently been directly substantiated by the demonstration that the *eve* protein can repress transcription *in vitro*[86].

Cell-type transcription factors, master regulators of cell type, and hierarchies. The proteins that are ultimately responsible for the differences between **a** and α cells are encoded by the mating-type locus, which is thus the master regulatory locus that determines yeast cell type[87]. Yeast has been a particularly favourable organism for identifying such a determinant: because yeast cells naturally differ at the mating-type locus[88], it was apparent that the alleles of this locus are the determinants for cell type[4,89]. Many other organisms, in particular, fungi and algae, have natural alleles governing their life cycle which may also identify master regulatory loci for these organisms (reviewed in ref. 1). The route to identifying transcription factors responsible for cell specialization in multicellular organisms—potential master regulatory loci—is coming from several different avenues: from isolation of mutants, from identification of DNA-binding proteins, and from use of DNA transfection (reviewed in ref. 90). The *glass* gene, identified genetically, is a putative cell-type transcription factor necessary for development of the *Drosophila* photoreceptor[91]. The Pit-1 protein, identified by its binding to the upstream regions of mammalian prolactin and growth-hormone genes, may be responsible for pituitary-specific expression of these genes[34,92,93]. A group of genes (the products of which are MyoD1, myogenin and Myf-5), cloned by transfection and subsequent cross-hybridization[94–97], seem to be regulatory proteins that govern myogenesis in mammalian cells.

It may be instructive to view the transfection technique from the perspective of yeast cell specialization. The key feature of the transfection strategy is that it identifies a DNA segment (genomic or complementary DNA) that stimulates a recipient cell to undergo an identifiable change, such as production of proteins characteristic of a differentiated cell. The factor responsible for this behaviour must be dominant in its action, and it ultimately must activate transcription in the recipient cell. The key point is that this added factor is sufficient to trigger cell differentiation. The transfectant DNA might code for a transcriptional activator protein that is absent from the recipient cell. This would be analogous to adding α1 protein to **a** cells to activate transcription of the α-specific genes in these cells. Another possibility is that the transfectant DNA codes for a transcriptional repressor protein, which is lacking in the recipient cell, that blocks synthesis of a negative factor. This would be analogous to adding the α2 protein to an **a** cell, to generate **a**1-α2, which would repress synthesis of *RME1* and thereby allow expression of the sporulation-specific gene set (see Fig. 1).

The success of the transfection technique is critically dependent on finding a suitable recipient cell line[90]. Indeed, the MyoD1 protein can activate muscle-specific gene expression in a wide variety of cell types but not in all cell types[98]. We can see precisely this behaviour if we consider the consequences of transferring the α1 gene into two different recipients[99]. Synthesis of α1 in **a** cells leads to expression of the α-specific genes: the only thing lacking for expression of these genes is α1. By contrast, synthesis of α1 in **a**/α cells does not lead to expression of α-specific genes: the presence of α1 and MCM1 is not sufficient to activate transcription. This failure is a consequence of two kinds of repression exerted by **a**1-α2. First, **a**1 − α2 represses synthesis of STE12 (ref. 100) and STE5 (J. Thorner, personal communication), both of which are necessary for maximal transcription of the α-specific genes[57,58]. Second, at least some of the upstream regions of α-specific genes contain sites for **a**1-α2 (ref. 101). We thus see that there are many reasons why **a**/α cells do not transcribe the α-specific genes: they lack at least three proteins necessary for transcription of these genes, and they may also have a repressor bound to their upstream regions.

In the simplest view of things, MyoD1 is analogous to α1: its presence is sufficient to activate transcription, and it stands at the top of the hierarchy for cell specialization, where, like α1, it directly stimulates transcription of the members of the specialized gene set. Of course, less direct roles are also possible: MyoD1 might activate synthesis of other proteins that directly promote transcription of this gene set. In any event, the concept[98]

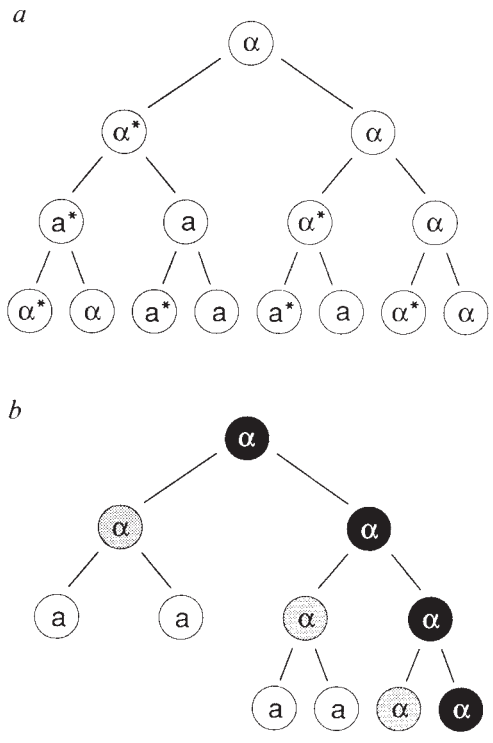

FIG. 6 Regulation of the cell-type regulatory genes generates a stem cell lineage. Panel *a* shows a mitotic cell lineage of yeast cells that carry the *HO* gene[71]. Mother cells are drawn to the left and daughters to the right at each cell division. Cells with an asterisk (cells that have undergone one cell-division cycle) are competent to produce cells with changed mating type in their next cell division[71]. This pattern results from two types of regulation of *HO*: it is transcribed only in mother cells and only in the G1 phase of the cell-division cycle (see text). Panel *b* accents analogies between the yeast cell lineage and a stem-cell/differentiated-cell lineage. The lineage of daughters (drawn in black) is a stem cell lineage; they give rise to cells like themselves at each cell division and also to cells (depicted as stippled) which are capable of giving rise to differentiated cells (drawn in white). In the yeast cell lineage, the 'white' cells can switch back to α; however, if this cell lineage were conducted in the presence of α-factor, the 'differentiated' **a** cells would be terminal.

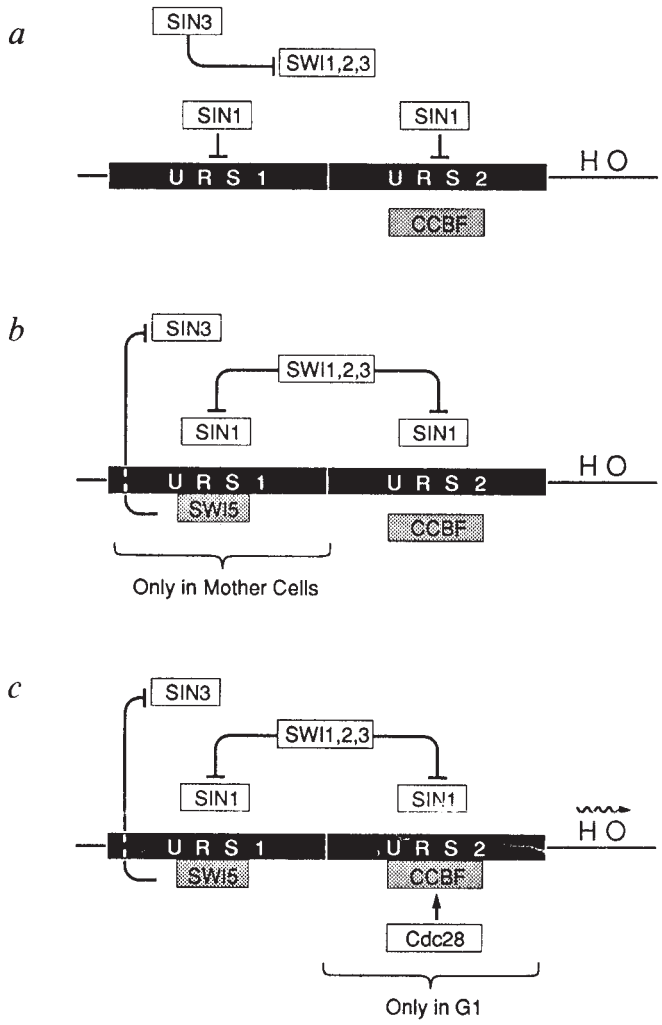

FIG. 7 Sequential activation scheme for expression of *HO*. The upstream regulatory region of *HO* contains two broad regions, URS1, which is responsible for mother-daughter regulation, and URS2, which is responsible for cell-cycle-dependent expression[73]. URS2 contains ten short sequences (cell-cycle boxes; not shown) that confer cell-cycle-dependent expression and that are binding sites for CCBF (cell-cycle box factor). It is proposed that two independent events must occur to activate transcription of the *HO* gene—one event in URS1 (in mother cells) and another event in URS2 (in G1). Arrowheads indicate that a gene product stimulates a process or is necessary for a subsequent event; bars indicate that a gene product inhibits a process or another product. Inhibition of SWI1, 2, 3 by SIN3 and inhibition of SIN3 by SWI5 may occur by competition for binding sites in URS1. Only SWI5 and CCBF (drawn as shaded rectangles) are known to bind to DNA. Panel *a* shows the upstream regulatory region in a daughter cell, in which it is proposed that two negative factors (SIN1 and SIN3) keep the *HO* gene silent: SIN1 prevents access of CCBF to its target sites, the CCB elements located in URS2. SIN3 inhibits SWI1, 2, 3 until the appropriate conditions occur. Panels *b* and *c* show the series of events that is proposed to occur in mother cells according to the sequential activation scheme (B. Andrews, W. Kruger, C. Peterson and I. H., in preparation). First, SWI5 protein inhibits SIN3 (an event that occurs only in mother cells). As a result, the second step can occur: SWI1, 2, 3 inhibits SIN1 protein. Now that repression by SIN1 is lifted, the third and final event can take place (*c*): binding of CCBF to its target sites, the cell-cycle boxes. Activation of transcription by CCBF occurs only in the G1 phase of the cell cycle and requires the CDC28 protein kinase, which may stimulate activity of CCBF (as drawn) or phosphorylate other components of the transcription machinery. A major motivation for the scheme depicted here is that mutations that inactivate the *SIN1* and *SIN2* genes bypass the need for *SWI1, SWI2, SWI3,* and *SWI5* genes for transcription of *HO*, whereas *SWI4* (and presumably *SWI6*) are still required[74]. The activity described as SIN1 may require also the *SIN2* gene; the activity described as SIN3 may require also the *SIN4* gene[74].

that MyoD1 represents a 'nodal point'—a point of potential regulation—in the pathway of muscle cell differentiation is a valuable one. Although it is clear that MyoD1 and relatives are sufficient to induce myogenesis, it is not known whether they are necessary and if they play this role in the whole organism. This will require inactivating the *MyoD1* gene, for example, in transgenic mice by gene replacement or by using dominant negative versions of MyoD1 (ref. 102). We may pursue the analogy between MyoD1 and α1 further and ask whether there is a mammalian analogue of MCM1: that is, whether MyoD1 itself activates transcription or whether it helps another protein to do so.

The regulatory hierarchy that is responsible for yeast cell type is crowned by sequence-specific DNA-binding proteins. Although DNA-protein recognition is one mechanism for generating specificity at a molecular level, and is used extensively for programming early *Drosophila* development (see, for example, ref. 103; reviewed in ref. 64), it is by no means the only mechanism. Indeed, the yeast hierarchy demonstrates the use of genetic rearrangement to control the choice of regulatory proteins. And the genetic hierarchy used for sex determination by *Drosophila* makes extensive use of alternative splicing to generate male- and female-specific pathways of development[104]. The steps in a regulatory hierarchy may use a wide variety of molecular mechanisms, including some not described here (such as proteolysis; see ref. 87).

Devices for monitoring progress and state. The process of development in multicellular organisms must include ways of monitoring the status of a cell and whether specific events have been executed. One of the important strategies for monitoring is the use of negative feedback control, whereby end products of a pathway signal their presence by inhibiting their further synthesis. The heteromeric repressor species, **a**1-α2, is used as such a monitoring device in several different respects. First of all, **a**1-α2 provides a molecular signal that successful mating to produce a diploid cell has been accomplished. Diploidy *per se* is not monitored: diploid **a**/**a** or α/α cells (produced by mitotic recombination from **a**/α cells) behave as their respective haploids in that they produce mating factors and so on. It is **a**1-α2, composed of products from both mating partners, that gives the **a**/α cell its identity. Repression of *HO* by **a**1-α2 is another good example of negative feedback regulation, in this case, a process that involves genetic rearrangement (see refs. 1, 11). As noted above, positive feedback also plays an important role in the mating process, whereby partially differentiated cells induce each other's final differentiation. It is intriguing that the upstream regulatory region of the *STE12* gene contains several sites for induction[53,54], which might serve either for positive or negative feedback.

One of the strategies that is used repeatedly for monitoring status and progress in yeast is that of combinatorial control. There are different types and levels of such control, and we have already seen one: the formation of a novel regulatory species, **a**1-α2, by combining polypeptides. Combinatorial control is also responsible for restricting mating-type switching, not only to certain cell types, but also to when it occurs in the cell cycle and in which cells it occurs. Three independent conditions must be satisfied before the *HO* gene is expressed. Looking at the *HO* upstream regulatory region (at present primarily using our genetic eyeglasses), we can see the molecular basis for this

combinatorial control. In this case, we think that events in one segment of DNA are coupled to events in another DNA segment in a dependent manner.

Violins and orchestras. Because yeast is a single-cell organism, one might at first have thought that its molecular and cellular sophistication would be modest. However, yeast cells do a lot more than simply double every 100 minutes: they can also mate and sporulate. It is the cells that are specialized for these life-cycle transitions which exhibit such interesting properties and which potentially shed light on differentiation in multicellular eukaryotes. Yeast presents an integrated picture of how regulatory proteins and intercellular signalling fit together to determine cell type. In trying to unravel the logical and molecular basis of development in multicellular organisms, it may be helpful to consider what a single cell can do. □

Ira Herskowitz is in the Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA.

1. Herskowitz, I. *Microbiol. Rev.* **52,** 536–553 (1988).
2. Cross, F., Hartwell, L. H., Jackson, C. & Konopka, J. B. *A. Rev. Cell Biol.* **4,** 429–57 (1988).
3. Nasmyth, K. & Shore, D. *Science* **237,** 1162–1170 (1987).
4. Strathern, J., Hicks, J. & Herskowitz, I. *J. molec. Biol.* **147,** 357–372 (1981).
5. Astell, C. R. *et al. Cell* **27,** 15–23 (1981).
6. Goutte, C. & Johnson, A. D. *Cell* **52,** 875–882 (1988).
7. Klar, A. J. S., Strathern, J. N., Broach, J. R. & Hicks, J. B. *Nature* **289,** 239–244 (1981).
8. Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C. & Smith, M. *Nature* **289,** 244–250 (1981).
9. Sprague, G. F. Jr., Jensen, R. & Herskowitz, I. *Cell* **32,** 409–415 (1983).
10. Wilson, K. L. & Herskowitz, I. *Molec. cell. Biol.* **4,** 2420–2427 (1984).
11. Jensen, R., Sprague, G. F. Jr. & Herskowitz, I. *Proc. natn. Acad. Sci. U.S.A.* **80,** 3035–3039 (1983).
12. Mitcheli, A. P. & Herskowitz, I. *Nature* **319,** 738–742 (1986).
13. Jarvis, E. E., Hagen, D. C. & G. F. Sprague, Jr. *Molec. cell. Biol.* **8,** 309–320 (1988).
14. Inokuchi, K., Nakayama, A. & Hishinuma, F. *Molec. cell. Biol.* **7,** 3185–3193 (1987).
15. Johnson, A. D. & Herskowitz, I. *Cell* **42,** 237–247 (1985).
16. Miller, A. M., MacKay, V. L. & Nasmyth, K. A. *Nature* **314,** 598–603 (1985).
17. Keleher, C. A., Goutte, C. & Johnson, A. D. *Cell* **53,** 927–936 (1988).
18. Bender, A. & Sprague, G. F. Jr *Cell* **50,** 681–691 (1987).
19. Tan, S., Ammerer, G. & Richmond, T. J. *EMBO J.* **7,** 4255–4264 (1988).
20. Kassir, Y. & Simchen, G. *Genetics* **82,** 187–206 (1976).
21. Kassir, Y., Granot, D. & Simchen, G. *Cell* **52,** 853–862 (1988).
22. Keleher, C. A., Passmore, S. & Johnson, A. D. *Molec. cell. Biol.* **9,** 5228–5230 (1989).
23. Passmore, S., Elble, R. & Tye, B.-K. *Genes Dev.* **3,** 921–935 (1989).
24. Jarvis, E. E., Clark, K. L. & Sprague, G. F. Jr *Genes Dev.* **3,** 936–945 (1989).
25. Kronstad, J. W., Holly, J. A. & MacKay, V. L. *Cell* **50,** 369–377 (1987).
26. Passmore, S., Maine, G. T., Elble, R., Christ, C. & Tye, B.-K. *J. molec. Biol.* **204,** 593–606 (1988).
27. Berry, M., Nunez, A. - M. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **86,** 1218–1222 (1989).
28. Ptashne, M. *Nature* **335,** 683–689 (1988).
29. Norman, C., Runswick, M., Pollock, R. & Treisman, R. *Cell* **55,** 989–1003 (1988).
30. Hayes, T. E., Sengupta, P. & Cochran, B. H. *Genes Dev.* **2,** 1713–1722 (1988).
31. Johnston, M. A. *Microbiol. Rev.* **51,** 458–476 (1987).
32. Shepard, J. C. W., McGinnis, W., Carrasco, A. E., De Robertis, E. M. & Gehring, W. J. *Nature* **310,** 70–71 (1984).
33. Laughon, A. & Scott, M. P. *Nature* **310,** 25–31 (1984).
34. He, X. *et al. Nature* **340,** 35–42 (1989).
35. Siliciano, P. G. & Tatchell, K. *Proc. natn. Acad. Sci. U.S.A.* **83,** 2320–2324 (1986).
36. Errede, B., Company, M. & Hutchison, C. A. III *Molec. cell. Biol.* **7,** 258–265 (1987).
37. Winston, F., *et al. Genetics* **115,** 649–656 (1987).
38. Stern, M., Jensen, R. E. & Herskowitz, I. *J. molec. Biol.* **178,** 853–868 (1984).
39. Trueheart, J., Boeke, J. D. & Fink, G. R. *Molec. cell. Biol.* **7,** 2316–2328 (1987).
40. McCaffrey, G., Clay, F. J., Kelsay, K. & Sprague, G. F. Jr *Molec. cell. Biol.* **7,** 2680–2690 (1987).
41. Hagen, D. C. & Sprague, G. F. Jr *J. molec. Biol.* **178,** 835–852 (1984).
42. Hartig, A., Holly, J., Saari, G. & MacKay, V. L. *Molec. cell. Biol.* **6,** 2106–2114 (1986).
43. Anderegg, R. J., Betz, R., Carr, S. A., Crabb, J. W. & Duntze, W. *J. biol. Chem.* **263,** 18236–18240 (1988).
44. Wilkinson, L. W. & Pringle, J. R. *Expl Cell Res.* **89,** 175–189 (1974).
45. Bücking-Throm, E., Duntze, W., Hartwell, L. H. & Manney, T. R. *Expl Cell Res.* **76,** 99–110 (1973).
46. Reed, S. I., Hadwiger, J. A. & Lorincz, A. T. *Proc. natn. Acad. Sci. U.S.A.* **82,** 4055–4059 (1985).
47. Mendenhall, M. D., Jones, C. A. & Reed, S. I. *Cell* **50,** 927–935 (1987).
48. Pringle, J. R. & Hartwell, L. H. In *The Molecular Biology of the Yeast Saccharomyces* (eds Strathern, J. N., Jones, E. W. & Brach, J. R.) 97–142 (Cold Spring Harbor Press, New York, 1981).
49. Marsh, L. & Herskowitz, I. *Cold Spring Harb. Symp. quant. Biol.* **53,** 557–565 (1988).
50. Dietzel, C. & Kurjan, J. *Cell* **50,** 1001–1010 (1987).
51. Van Arsdell, S. W., Stetler, G. L. & Thorner, J. *Molec. cell. Biol.* **7,** 749–759 (1987).
52. Teague, M. A., Chaleff, D. T. & Errede, B. *Proc. natn. Acad. Sci. U.S.A.* **83,** 7371–7375 (1986).
53. Dolan, J. W., Kirkman, C. & Fields, S. *Proc. natn. Acad. Sci. U.S.A.* **86,** 5703–5707 (1989).
54. Errede, B. & Ammerer, G. *Genes Dev.* **3,** 1349–1361 (1989).
55. Sorger, P. K. & Pelham, H. R. B. *Cell* **54,** 855–864 (1988).
56. Mylin, L. M., Bhat, J. P. & Hopper, J. E. *Genes Dev.* **3,** 1157–1165 (1989).
57. Fields, S. & Herskowitz, I. *Cell* **42,** 923–930 (1985).
58. Fields, S., Chaleff, D. T. & Sprague, G. F. Jr *Molec. cell. Biol.* **8,** 551–556 (1988).
59. Weeks, D. L. & Melton, D. A. *Cell* **51,** 861–867 (1987).
60. Tomlinson, A. & Ready, D. F. *Devl Biol.* **120,** 366–376 (1987).
61. Zipursky, S. L. *Trends Neurosci.* **12,** 183–189 (1989).
62. Sternberg, P. W. & Horvitz, H. R. *Cell* **58,** 679–693 (1989).
63. Strazdis, J. R. & MacKay, V. L. *Nature* **305,** 543–545 (1983).
64. Ingham, P. W. *Nature* **335,** 25–34 (1988).
65. Lewis, E. B. *Nature* **276,** 565–570 (1978).
66. Hicks, J. B., Strathern, J. N. & Herskowitz, I. in *DNA Insertion Elements, Plasmids and Episomes* (eds Bukhari, A., Shapiro, J. & Adhya, S.) 457–462 (Cold Spring Harbor Press, New York, 1977).
67. Herskowitz, I. & Oshima, Y. in *The Molecular Biology of the Yeast Saccharomyces cerevisiae: Life Cycle and Inheritance* (eds Strathern, J. N., Jones, E. W., & Broach, J. R.) 181–209 (Cold Spring Harbor Press, New York, 1981).
68. Kostriken, R. & Heffron, F. *Cold Spring Harb. Symp. quant. Biol.* **49,** 89–96 (1984).
69. Nasmyth, K. *Cell* **42,** 225–235 (1985).
70. Nasmyth, K. *Nature* **302,** 670–676 (1983).
71. Strathern, J. N. & Herskowitz, I. *Cell* **17,** 371–381 (1979).
72. Nasmyth, K. *Cell* **42,** 213–223 (1985).
73. Breeden, L. & Nasmyth, K. *Cell* **48,** 389–397 (1987).
74. Sternberg, P. W., Stern, M. J., Clark, I. & Herskowitz, I. *Cell* **48,** 567–577 (1987).
75. Nasmyth, K., Stillman, D. & Kipling, D. *Cell* **48,** 579–587 (1987).
76. Andrews, B. J. & Herskowitz, I. *Cell* **57,** 21–29 (1989).
77. Cisek, L. J. & Corden, J. L. *Nature* **339,** 679–684 (1989).
78. Stillman, D. J., Bankier, A. T., Seddon, A., Groenhout, E. G. & Nasmyth, K. A. *EMBO J.* **7,** 485–494 (1988).
79. Nagai, K., Nakesoko, Y., Nasmyth, K. & Rhodes, D. *Nature* **332,** 284–286 (1988).
80. Nasmyth, K., Seddon, A. & Ammerer, G. *Cell* **49,** 549–558 (1987).
81. Guarente, L., Lalonde, B., Gifford, P. & Alani, E. *Cell* **36,** 503–511 (1983).
82. Beach, D., Durkacz, B. & Nurse, P. *Nature* **300,** 706–709 (1982).
83. Gautier, J., Norbury, C., Lohka, M., Nurse, P. & Maller, J. *Cell* **54,** 433–439 (1988).
84. Breeden, L. & Nasmyth, K. *Nature* **329,** 651–654 (1987).
85. Yochem, J. & Greenwald, I. *Cell* **58,** 553–563 (1989).
86. Biggin, M. D. & Tjian, R. *Cell* **58,** 433–440 (1989).
87. Herskowitz, I. *Cold Spring Harb. Symp. quant. Biol.* **50,** 565–574 (1985).
88. Lindegren, C. C. & Lindegren, G. *Ann. Mo. bot. Gdn* **30,** 453–469 (1943).
89. MacKay, V. L. & Manney, T. R. *Genetics* **76,** 273–288 (1974).
90. Blau, H. M. *Cell* **53,** 673–674 (1988).
91. Moses, K., Ellis, M. C. & Rubin, G. M. *Nature* **340,** 531–536 (1989).
92. Bodner, M. *et al. Cell* **55,** 505–518 (1988).
93. Ingraham, H. A. *et al. Cell* **55,** 519–529 (1988).
94. Davis, R. L., Weintraub, H. & Lassar, A. B. *Cell* **51,** 987–1000 (1987).
95. Wright, W. E., Sassoon, D. A. & Lin, V. K. *Cell* **56,** 607–617 (1989).
96. Edmondson, D. G. & Olson, E. N. *Genes Dev.* **3,** 628–640 (1989).
97. Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E. & Arnold, H. H. *EMBO J.* **8,** 701–709 (1989).
98. Weintraub, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **86,** 5434–5438 (1989).
99. Ammerer, G., Sprague, G. F. Jr & Bender, A. *Proc. natn. Acad. Sci. U.S.A.* **82,** 5855–5859 (1985).
100. Fields, S. & Herskowitz, I. *Molec. cell. Biol.* **7,** 3818–3821 (1987).
101. Flessel, M. C., Brake, A. J. & Thorner, J. *Genetics* **121,** 223–236 (1989).
102. Herskowitz, I. *Nature* **329,** 219–222 (1987).
103. Driever, W. & Nüsslein-Volhard, C. *Nature* **337,** 138–143 (1989).
104. Baker, B. S. *Nature* **340,** 521–524 (1989).
105. McGrath, J. P. & Varshavsky, A. *Nature* **340,** 400–404 (1989).
106. Elder, R. T., St. John, T. P., Stinchcomb, D. T. & Davis, R. W. *Cold Spring Harb. Symp. quant. Biol.* **45,** 581–584 (1981).
107. Errede, B., Cardillo, T. S., Wever, G. & Sherman, F. *Cold Spring Harb. Symp. quant. Biol.* **45,** 593–602 (1981).
108. Whiteway, M. *et al. Cell* **56,** 467–477 (1989).
109. Dranginis, A. *Molec. cell. Biol.* **9,** 3992–3998 (1989).
110. Miyajima, I. *et al. Cell* **50,** 1011–1019 (1987).
111. Jahng, K.-Y., Ferguson, J. & Reed, S. I. *Molec. cell. Biol.* **8,** 1440–1448 (1988).
112. Ivy, J. M., Klar, A. J. S. & Hicks, J. B. *Molec. cell. Biol.* **6,** 688–702 (1986).
113. Strathern, J. N. *et al. Cell* **31,** 183–192 (1982).
114. Sauer, R. T., Smith, D. L. & Johnson, A. D. *Genes Dev.* **2,** 1713–1722 (1988).
115. Andrews, B. J. & Herskowitz, I. *Nature* (in the press).

ACKNOWLEDGEMENTS. I thank Sandy Johnson, Rudi Grosschedl, Flora Banuett, Brenda Andrews, Craig Peterson and Paul Sternberg for comments on the manuscript, Bob Horvitz, Beverly Errede, Barbara Wold and Cindy Keleherv for discussion, Beverly Errede, Carolyn Goutte, Sandy Johnson and the members of my laboratory for allowing me to cite work before publication, and Kerriew Andow for preparing the figures. Work from my laboratory has been supported by grants from the US National Institutes of Health and from the Weingart Foundation.