REARRANGEMENT OF DNA

4.18.1 Introduction

Although genomic DNA is usually unaltered by somatic development, there are some cases in which sequences are moved within a genome, modified, amplified, or even lost, as a natural event. In this chapter, we discuss a variety of such events in yeast, plants, and lower eukaryotes. Examples of rearrangement or loss of specific sequences are especially common in the lower eukaryotes. Usually these changes involve somatic cells; the germline remains inviolate. (However, there are organisms whose reproductive cycle involves the loss of whole chromosomes or sets of chromosomes.) We also discuss the introduction of new sequences into the genome by experimental means. Reorganization of particular sequences is rare in animals, although an extensive case is represented by the immune system. In *Molecular Biology 5.25 Immune diversity*, we discuss the rearrangement and expression of immunoglobulin genes.

There are two types of circumstances in which gene rearrangement is used to control expression:

- Rearrangement may create new genes, needed for expression in particular circumstances, as in the case of the immunoglobulins.
- Rearrangement may be responsible for switching expression from one preexisting gene to another. This provides a mechanism for regulating gene expression.

Yeast mating type switching and trypanosome antigen variation share a similar type of plan in which gene expression is controlled by manipulation of DNA sequences. Phenotype is determined by the gene copy present at a particular, active locus. But the genome also contains a store of other, alternative sequences, which are silent. A silent copy can be activated only by a rearrangement of sequences in which it replaces the active gene copy. Such a substitution is equivalent to a unidirectional transposition with a specific target site.

The simplest example of this strategy is found in the yeast, *S. cerevisiae*. Haploid *S. cerevisiae* can have either of two mating types. The type is determined by the sequence present at the active mating type locus. But the genome also contains two other, silent loci, one representing each mating type. Transition between mating types is accomplished by substituting the sequence at the active locus with the sequence from the silent locus carrying the other mating type.

A range of variations is made possible by DNA rearrangement in the African trypanosomes, unicellular parasites that evade the host immune response by varying their surface antigens. The type of surface antigen is determined by the gene sequence at an active locus. This sequence can be changed, however, by substituting a sequence from any one of many silent loci. It seems fitting that the mechanism used to combat the flexibility of the immune apparatus is analogous to that used to generate immune diversity: it relies on physical rearrangements in the genome to

change the sequences that are expressed.

Another means of increasing genetic capacity is employed in parasite- or symbiote-host interactions, in which exogenous DNA is introduced from a bacterium into a host cell. The mechanism resembles bacterial conjugation. Expression of the bacterial DNA in its new host changes the phenotype of the cell. In the example of the bacterium *Agrobacterium tumefaciens*, the result is to induce tumor formation by an infected plant cell.

Alterations in the relative proportions of components of the genome during somatic development occur to allow insect larvae to increase the number of copies of certain genes. And the occasional amplification of genes in cultured mammalian cells is indicated by our ability to select variant cells with an increased copy number of some gene. Initiated within the genome, the amplification event can create additional copies of the gene that survive in either intrachromosomal or extrachromosomal form.

When extraneous DNA is introduced into eukaryotic cells, it may give rise to extrachromosomal forms or may be integrated into the genome. The relationship between the extrachromosomal and genomic forms is irregular, depending on chance and to some degree unpredictable events, rather than resembling the regular interchange between free and integrated forms of bacterial plasmids.

Yet, however accomplished, the process may lead to stable change in the genome; following its injection into animal eggs, DNA may even be incorporated into the genome and inherited thereafter as a normal component, sometimes continuing to function. Injected DNA may enter the germline as well as the soma, creating a transgenic animal. The ability to introduce specific genes that function in an appropriate manner could become a major medical technique for curing genetic diseases.

The converse of the introduction of new genes is the ability to disrupt specific endogenous genes. Additional DNA can be introduced within a gene to prevent its expression and to generate a null allele. Breeding from an animal with a null allele can generate a homozygous "knockout," which has no active copy of the gene. This is a powerful method to investigate directly the importance and function of a gene.

Considerable manipulation of DNA sequences therefore is achieved both in authentic situations and by experimental fiat. We are only just beginning to work out the mechanisms that permit the cell to respond to selective pressure by changing its bank of sequences or that allow it to accommodate the intrusion of additional sequences.

REARRANGEMENT OF DNA

4.18.2 The mating pathway is triggered by pheromone-receptor interactions

Key Terms

The **mating type** is a property of haploid yeast that makes it able to fuse to form a diploid only with a cell of the opposite mating type.

A **pheromone** is a small molecule secreted by one mating type of an organism in order to interact with a member of the opposite mating type.

Key Concepts

- Yeast of a given mating type secrete a small polypeptide that binds to a receptor on cells of the opposite mating type.

The yeast *S. cerevisiae* can propagate in either the haploid or diploid condition. Conversion between these states takes place by mating (fusion of haploid spores to give a diploid) and by sporulation (meiosis of diploids to give haploid spores). The ability to engage in these activities is determined by the **mating type** of the strain.

The properties of the two mating types are summarized in **Figure 18.1**. We may view them as resting on the rationale that there is no point in mating unless the haploids are of different genetic types; and sporulation is productive only when the diploid is heterozygous and therefore able to generate recombinants.

Haploids and diploids are different

	MATa	*MATα*	*MATa*/*MATα*
Cell type	**a**	α	**a**/α
Mating	yes	yes	no
Sporulation	no	no	yes
Pheromone	**a** factor	a factor	none
Receptor	binds α factor	binds **a** factor	none

Figure 18.1 Mating type controls several activities.

The mating type of a (haploid) cell is determined by the genetic information present at the MAT locus. Cells that carry the *MATa* allele at this locus are type *a*; likewise, cells that carry the MAT α allele are type α. Cells of opposite types can mate; cells of the same type cannot.

Recognition of cells of opposite mating type is accomplished by the secretion of **pheromones**. α cells secrete the small polypeptide α-factor; *a* cells secrete *a*-factor. The α-factor is a peptide of 13 amino acids; the *a*-factor is a peptide of 12 amino acids that is modified by addition of a farnesyl (lipid-like) group and carboxymethylation. Each of these peptides is synthesized in the form of a precursor

polypeptide that is cleaved to release the mature peptide sequence.

A cell of one mating type carries a surface receptor for the pheromone of the opposite type. When an *a* cell and an α cell encounter one another, the pheromones bind to their receptors to arrest the cells in the G1 phase of the cell cycle, and various morphological changes occur. In a successful mating, the cell cycle arrest is followed by cell and nuclear fusion to produce an *a*/ α diploid cell.

The *a*/ α cell carries both the *MATa* and *MAT* α alleles and has the ability to sporulate. **Figure 18.2** demonstrates how this design maintains the normal haploid/diploid life cycle. Note that only heterozygous diploids can sporulate; homozygous diploids (either *a/a* or α/ α) cannot sporulate.

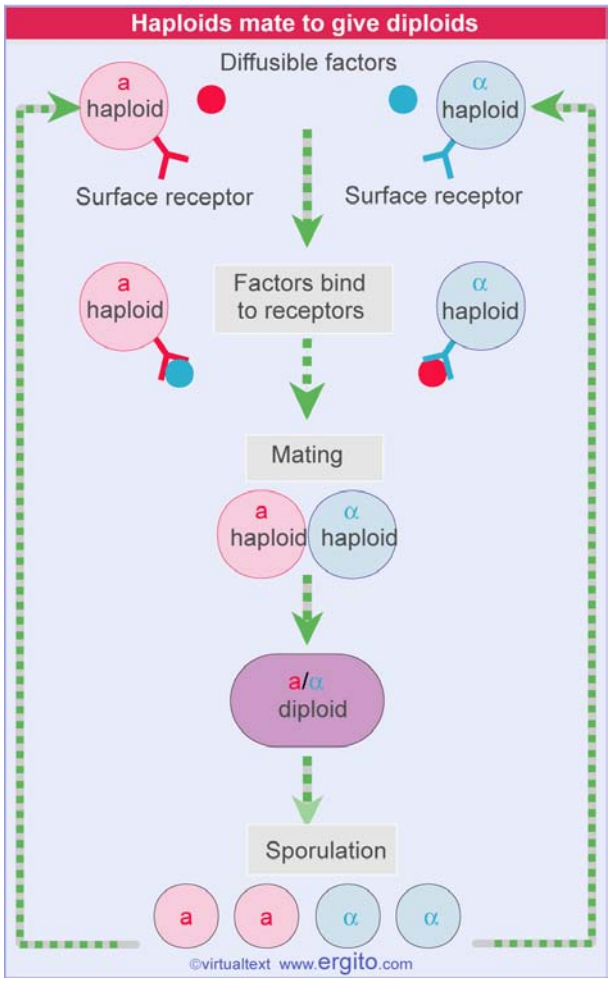

Figure 18.2 The yeast life cycle proceeds through mating of *MATa* and *MAT* α haploids to give heterozygous diploids that sporulate to generate haploid spores.

Much of the information about the yeast mating type pathway was deduced from the properties of mutations that eliminate the ability of *a* and/or α cells to mate. The genes identified by such mutations are called *STE* (for sterile). Mutations in the genes *STE2* and *STE3* are specific for individual mating types; but mutations in the

other *STE* genes eliminate mating in both *a* and α cells. This situation is explained by the fact that the events that follow the interaction of factor with receptor are identical for both types (for review see 172).

Mating is a symmetrical process that is initiated by the interaction of pheromone secreted by one cell type with the receptor carried by the other cell type. The only genes that are uniquely required for the response pathway in a particular mating type are those coding for the receptors. Either the *a* factor-receptor or the α factor-receptor interaction switches on the same response pathway. Mutations that eliminate steps in the common pathway have the same effects in both cell types (584).

Reviews

172. Nasmyth, K. (1982). *Molecular genetics of yeast mating type.* Annu. Rev. Genet. 16, 439-500.

References

584. Bender, A. and Sprague, G. F., Jr. (1986). *Yeast peptide pheromones, a-factor and alpha-factor, activate a common response mechanism in their target cells.* Cell 47, 929-937.

REARRANGEMENT OF DNA

4.18.3 The mating response activates a G protein

Key Concepts

- Binding of mating type factor to receptor activates a trimeric G protein.
- The β γ dimer is released and activates a signal transduction pathway.

The initial steps in the mating-type response are summarized in **Figure 18.3**. The components are similar to those of the "classical" receptor-G protein coupled systems, in which a membrane receptor interacts with a trimeric G protein (see *Molecular Biology 6.28.5 G proteins may activate or inhibit target proteins*). Ste2 is the α-receptor in the *a* cell; Ste3 is the *a*-receptor in the α cell. When either receptor is activated, it interacts with the same G protein. *This means that the identical pathway is triggered when either type of mating factor interacts with a receptor of the opposite type* (584).

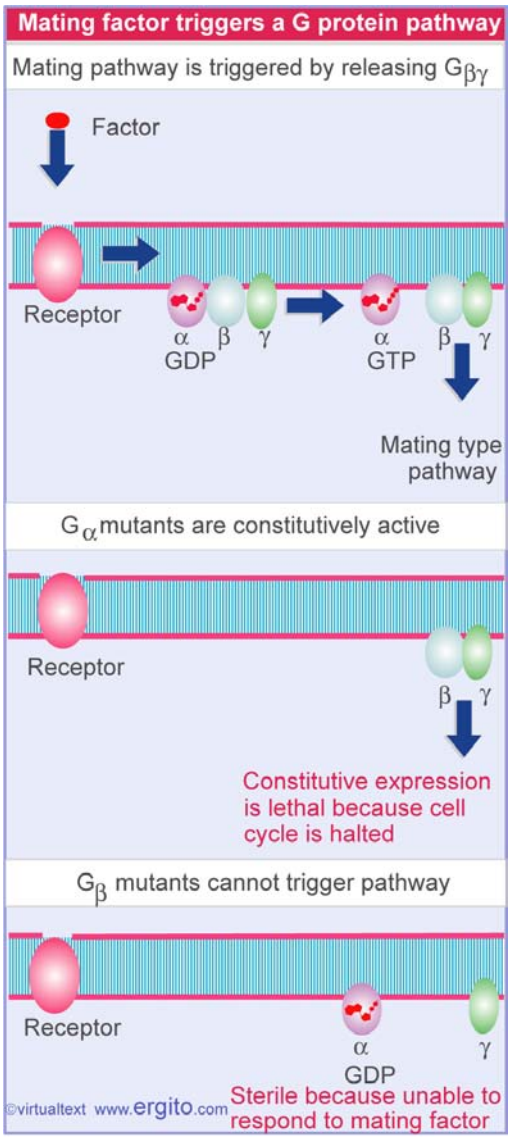

Figure 18.3 Either *a* or α factor/receptor interaction triggers the activation of a G protein, whose β γ subunits transduce the signal to the next stage in the pathway.

The trimeric G protein consists of the subunits, α, β, and γ. The α subunit binds a guanine nucleotide. In the intact (trimeric) G protein, the α subunit carries GDP. When the pheromone receptor is activated, it causes the GDP to be displaced by GTP. As a result, the α subunit is released from the β γ dimer. This separation of subunits allows the G protein to activate the next protein in the pathway (see *Molecular Biology 6.28.6 G proteins function by dissociation of the trimer*; for review see 2286).

The most common mechanism used in such pathways is for the activated α subunit to interact with the target protein. However, the situation is different in the mating type pathway, where the β γ dimer activates the next stage in the pathway. The component proteins of the G-trimer are identified by mutations in three genes, *SCG1*,

STE4, and *STE18*, that affect the response to binding pheromone. Inactivation of *SCG1*, which codes for the G_α protein, causes constitutive expression of the pheromone response pathway (because G_α is unable to maintain $G_{\beta\gamma}$ in the inactive trimeric form). The mutation is lethal, because its effects include arrest of the cell cycle. Inactivation of *STE4* (codes for G_β) or of *STE18* (codes for G_α) create sterility by abolishing the mating-type response (because the next step in the pathway cannot be activated; for review see 174; 176).

Reviews

174. Kurjan, J. (1992). *Pheromone response in yeast.* Annu. Rev. Biochem. 61, 1097-1129.

176. Kurjan, J. (1993). *The pheromone response pathway in S. cerevisiae.* Annu. Rev. Genet. 27, 147-179.

2286. Dohlman, H. G. and Thorner, J. W. (2001). *Regulation of G protein-initiated signal transduction in yeast: paradigms and principles.* Annu. Rev. Biochem. 70, 703-754.

References

584. Bender, A. and Sprague, G. F., Jr. (1986). *Yeast peptide pheromones, a-factor and alpha-factor, activate a common response mechanism in their target cells*. Cell 47, 929-937.

REARRANGEMENT OF DNA

4.18.4 The signal is passed to a kinase cascade

Key Concepts

- $G_{\beta\gamma}$ activates the monomeric G protein Cdc42.
- Cdc42 directly controls the structure of the cytoskeleton and activates a kinase cascade.
- In the kinase cascade, the signal passes through a series of kinases.
- The last kinase in the cascade activates a transcription factor and also phosphorylates other targets.
- The effect of the pathway is to repress functions needed for the cell cycle and to activate functions needed for mating.

Figure 18.4 summarizes the main steps of the mating type pathway (for review see 2286). (There are also some branches that are not shown in the figure.) When the $G_{\beta\gamma}$ dimer is released, it causes a monomeric G protein to be activated. Actually, the immediate target of $G_{\beta\gamma}$ is Cdc24, which is a nucleotide exchange factor that then activates the monomeric G protein Cdc42. The effect of Cdc24 on Cdc42 is a typical interaction in which the monomeric G protein is activated by replacing its bound GDP with GTP (see *Molecular Biology Supplement 32.10 G proteins*). Cdc42 resembles the Rho family of G proteins that control actin filament structures in higher eukaryotic cells (see *Molecular Biology 6.28.15 The activation of Ras is controlled by GTP*).

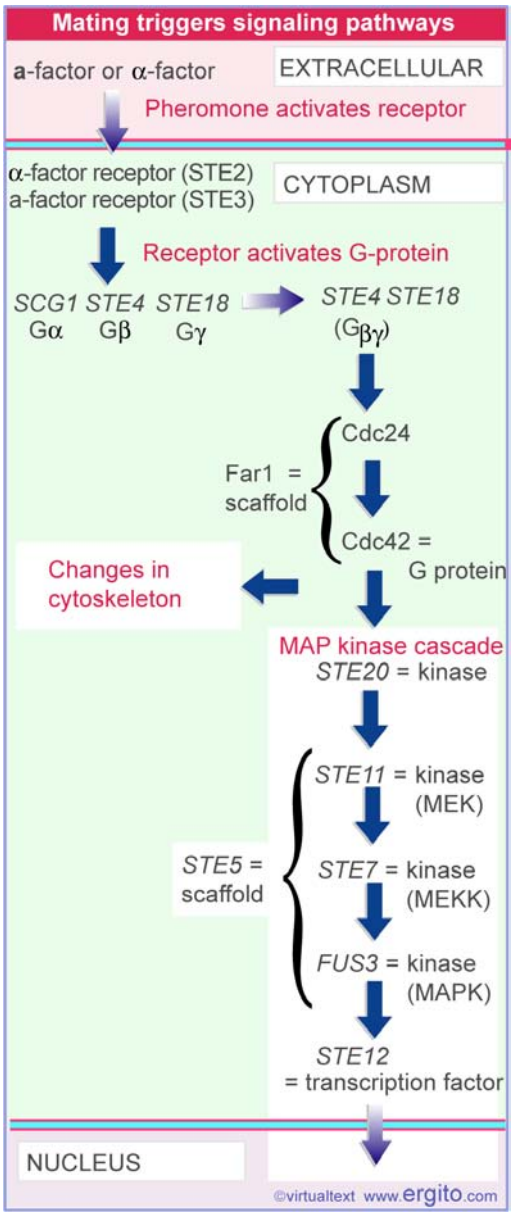

Figure 18.4 The same mating type response is triggered by interaction of either pheromone with its receptor. The signal is transmitted through a series of kinases to a transcription factor; there may be branches to some of the final functions.

Cdc42 then activates two pathways:

- one affects the structure of the cell by changing the organization of the cytoskeleton;
- the other is a cascade of kinases that ultimately activates transcription.

Cdc42 acts on many proteins that are involved in modifying the state of assembly of

actin filaments, and in this way changes the structure of the cell. It is necessary for budding, which generates the new daughter cell when division occurs. After mating, division stops and budding ceases, although cell growth continues in the direction of the pheromone. It is Cdc42 that is responsible for this change.

The pathway to regulating transcription is identified by a group of *STE* genes. They form a kinase cascade in which each member activates the next in the pathway by phosphorylating it. This is an example of a MAP kinase cascade (see *Molecular Biology 6.28.16 A MAP kinase pathway is a cascade*). Analogous pathways are found in higher organisms, and are compared with the yeast cascade in **Figure 28.38**. In the case of the yeast pathway, Cdc42 directly activates the kinase STE20. STE20 initiates the kinase cascade by activating STE11, which activates STE7, which activates FUS3, which finally triggers the response by phosphorylating the transcription factor STE12. This causes STE12 to migrate to the nucleus, where it activates transcription.

Some components in the signaling pathways do not have catalytic activities but act to assist the other components. Both FAR1 and STE5 in the pathways detailed in **Figure 18.4** are scaffold proteins that hold together the other members of the pathway. This may be necessary to ensure specificity in the response, for example, to ensure that each kinase in the MAPK cascade is directed to phosphorylate only the intended next member of the pathway (see *Molecular Biology 6.28.17 What determines specificity in signaling?*). Packaging the members of the pathway together also increases the efficiency of the response.

Both FAR1 and STE5 shuttle between the nucleus and cytoplasm. **Figure 18.5** shows that in a dividing cell, before the mating response is triggered, they are concentrated in the nucleus. **Figure 18.6** shows that they both interact with the activated $G_{\beta\gamma}$ dimer, and therefore become concentrated in the cytoplasm as a result of the pheromone-receptor interaction (2334). FAR1 is bound to Cdc24 (2333); the result of its relocalization to the cytoplasm is to enable the Cdc24 to activate Cdc42. STE5 binds to the three components of the MAPK cascade (808); actually, its interaction with the kinases may enhance the ability of STE20 to make the initial phosphorylation of STE11. So we see that, although $G_{\beta\gamma}$ directly activates the effector pathways through its effect on the Cdc24/Cdc42 interaction, it indirectly enhances the pathways by causing the necessary components to accumulate in the cytoplasm.

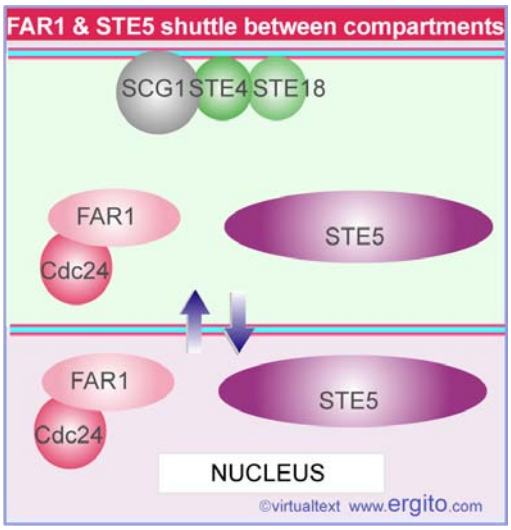

Figure 18.5 In a dividing cell, FAR1 (bound to Cdc24) and STE5 shuttle between the nucleus and cytoplasm. The $G_{\alpha\beta\gamma}$ complex of SCG1, STE4, STE18 is associated with the plasma membrane.

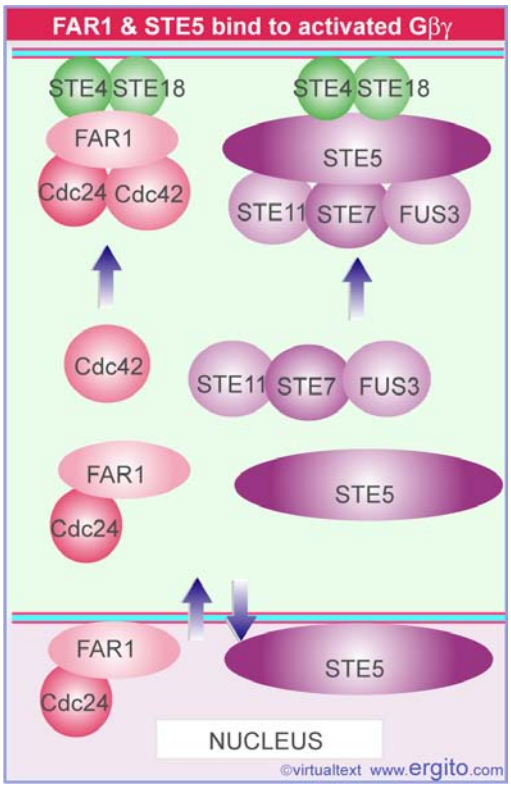

Figure 18.6 The mating pathway frees the active $G_{\beta\gamma}$ dimer (STE4-STE18). It forms complexes with the scaffolding proteins FAR1 and STE5. FAR1 is bound to Cdc24, which binds Cdc42. STE5 binds the three kinases of the MAPK pathway.

Branches from the cascade generate additional reactions. STE7 activates two kinases.

One is FUS3, whose action is dedicated to carrying forward the mating type pathway, as shown in **Figure 18.4**. FUS3 acts on earlier components of the pathway, specifically FAR1 and STE5, to enhance their actions, on the transcription factor STE12, and also on proteins that would otherwise inhibit STE12. The second target for STE7 is KSS1, which has ancillary functions (its main function is in vegetative growth).

Some of the end targets for the cascade are direct substrates of one of the kinases; for example, FUS3 kinase acts on Cln3, which is one of 3 Cln proteins needed for cell cycle progression. Other targets are controlled at the level of gene expression.

Last updated on 2-11-2002

Reviews

2286. Dohlman, H. G. and Thorner, J. W. (2001). *Regulation of G protein-initiated signal transduction in yeast: paradigms and principles.* Annu. Rev. Biochem. 70, 703-754.

References

808. Choi, K.-Y. et al. (1994). *Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in S. cerevisiae.* Cell 78, 499-512.

2333. Butty, A. C., Pryciak, P. M., Huang, L. S., Herskowitz, I., and Peter, M. (1998). *The role of Far1p in linking the heterotrimeric G protein to polarity establishment proteins during yeast mating.* Science 282, 1511-1516.

2334. Whiteway, M. S., Wu, C., Leeuw, T., Clark, K., Fourest-Lieuvin, A., Thomas, D. Y., and Leberer, E. (1995). *Association of the yeast pheromone response G protein beta gamma subunits with the MAP kinase scaffold Ste5p.* Science 269, 1572-1575.

REARRANGEMENT OF DNA

4.18.5 Yeast can switch silent and active loci for mating type

Key Terms

The **cassette model** for yeast mating type proposes that there is a single active locus (the active cassette) and two inactive copies of the locus (the silent cassettes). Mating type is changed when an active cassette of one type is replaced by a silent cassette of the other type.

Key Concepts

- The yeast mating type locus MAT has either the *MATa* or *MAT* α genotype.
- Yeast with the dominant allele HO switch their mating type at a frequency ~10^{-6}.
- The allele at *MAT* is called the active cassette.
- There are also two silent cassettes, HML α and HMR*a*.
- Switching occurs if *MATa* is replaced by HML α or *MAT* α is replaced by HMR*a*.

Yeast mating type is determined by the locus *MAT*, which can have either of the alleles *MATa* or *MAT* α. Some yeast strains have the remarkable ability to switch their mating types. These strains carry a dominant allele *HO* and change their mating type frequently, as often as once every generation. Strains with the recessive allele *ho* have a stable mating type, subject to change with a frequency ~10^{-6}.

The presence of *HO* causes the genotype of a yeast population to change. Irrespective of the initial mating type, in a very few generations there are large numbers of cells of both mating types, leading to the formation of *MATa/MAT* α diploids that take over the population. The production of stable diploids from a haploid population can be viewed as the raison d'être for switching.

The existence of switching suggests that all cells contain the potential information needed to be either *MATa* or *MAT* α, but express only one type. Where does the information to change mating types come from? Two additional loci are needed for switching. *HML* α is needed for switching to give a *MAT* α type; *HMRa* is needed for switching to give a *MATa* type. These loci lie on the same chromosome that carries *MAT*. *HML* is far to the left, *HMR* far to the right.

The **cassette** model for mating type is illustrated in **Figure 18.7**. It proposes that *MAT* has an *active cassette* of either type α or type *a*. *HML* and *HMR* have *silent cassette*s. Usually *HML* carries an α cassette, while *HMR* carries an *a* cassette. All cassettes carry information that codes for mating type, but only the active cassette at *MAT* is expressed. Mating-type switching occurs when the active cassette is replaced by information from a silent cassette. The newly installed cassette is then expressed (582).

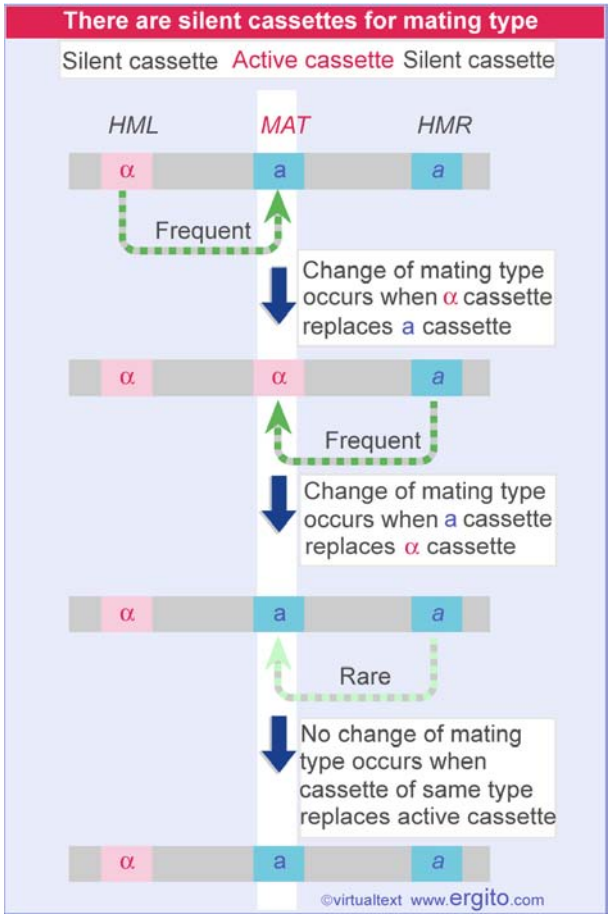

Figure 18.7 Changes of mating type occur when silent cassettes replace active cassettes of opposite genotype; when transpositions occur between cassettes of the same type, the mating type remains unaltered.

Switching is nonreciprocal; the copy at *HML* or *HMR* replaces the allele at *MAT*. We know this because a mutation at *MAT* is lost permanently when it is replaced by switching – it does not exchange with the copy that replaces it.

If the silent copy present at *HML* or *HMR* is mutated, switching introduces a mutant allele into the *MAT* locus. The mutant copy at *HML* or *HMR* remains there through an indefinite number of switches. Like replicative transposition, the donor element generates a new copy at the recipient site, while itself remaining inviolate.

Mating-type switching is a directed event, in which there is only one recipient (*MAT*), but two potential donors (*HML* and *HMR*). Switching usually involves replacement of *MATa* by the copy at *HML* α or replacement of *MAT* α by the copy at *HMRa*. In 80-90% of switches, the *MAT* allele is replaced by one of opposite type. This is determined by the phenotype of the cell. Cells of *a* phenotype preferentially choose *HML* as donor; cells of α phenotype preferentially choose *HMR*.

Several groups of genes are involved in establishing and switching mating type. As well as the genes that directly determine mating type, they include genes needed to

repress the silent cassettes, to switch mating type, or to execute the functions involved in mating.

By comparing the sequences of the two silent cassettes (*HML* α and *HMRa*) with the sequences of the two types of active cassette (*MATa* and *MAT* α), we can delineate the sequences that determine mating type. The organization of the mating type loci is summarized in **Figure 18.8**. Each cassette contains common sequences that flank a central region that differs in the *a* and α types of cassette (called *Ya* or *Y* α). On either side of this region, the flanking sequences are virtually identical, although they are shorter at *HMR*. The active cassette at *MAT* is transcribed from a promoter within the *Y* region.

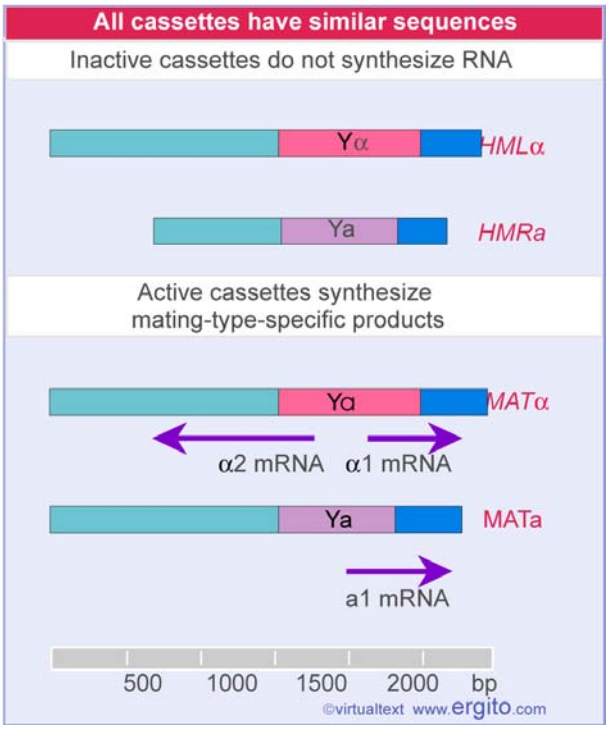

Figure 18.8 Silent cassettes have the same sequences as the corresponding active cassettes, except for the absence of the extreme flanking sequences in *HMRa*. Only the Y region changes between *a* and α types.

References

582. Hicks, J., Strathern, J. N., and Herskowitz, I. (1977). *The cassette model of mating type interconversion*. DNA Insertion Elements, 457-462.

REARRANGEMENT OF DNA

4.18.6 The *MAT* locus codes for regulator proteins

Key Concepts

- In α-type haploids, *MAT* α turns on genes required to specify α-specific functions required for mating, and turns off genes required for *a*-mating type.
- In *a*-type cells, *MATa* is not required.
- In diploids, *a1* and α2 products cooperate to repress haploid-specific genes.

The basic function of the *MAT* locus is to control expression of pheromone and receptor genes, and other functions involved in mating. *MAT* α codes for two proteins, α1 and α2. *MATa* codes for a single protein, *a1*. The *a* and α proteins directly control transcription of various target genes; they function by both positive and negative regulation. They function independently in haploids, and in conjunction in diploids. Their interactions are summarized in the table on the right of **Figure 18.9** in terms of three groups of target genes:

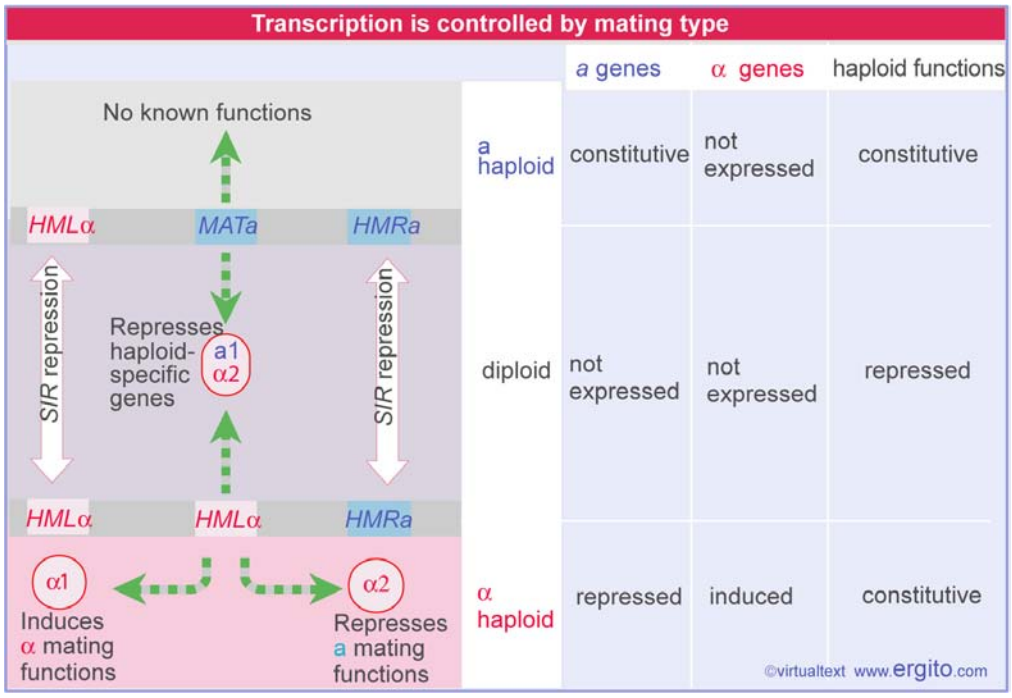

Figure 18.9 In diploids the α 1 and α 2 proteins cooperate to repress haploid-specific functions. In *a* haploids, mating functions are constitutive. In α haploids, the α 2 protein represses *a* mating functions, while α 1 induces α mating functions.

- *a*-specific genes are expressed constitutively in *a* cells. They are repressed in α cells. The *a*-specific genes include the *a*-factor structural gene, and *STE2*, which

codes for the α-factor receptor. So the *a* phenotype is associated with readiness to recognize the pheromone produced by the opposite mating type.

- α-specific functions are induced in α cells, but are not expressed in *a* cells. They include the α-factor structural gene, and the *a*-factor receptor gene, *STE3*. Again, the expression of pheromone of one type is associated with expression of receptor for the pheromone of the opposite type.

- Haploid-specific functions include genes that are needed for transcription of pheromone and receptor genes, the *HO* gene involved in switching, and *RME*, a repressor of sporulation. They are expressed constitutively in both types of haploid, but are repressed in a/ α diploids. As a result, the *a*-specific and α-specific functions also remain unexpressed in diploids.

We may now view the functions of the regulators and their targets from the perspective of the *MAT* functions expressed in haploid and diploid yeast cells, as outlined in the diagram on the left of **Figure 18.9**. The *a* and α mating types are regulated by different mechanisms:

- In *a* haploids, a mating functions are expressed constitutively. The functions of the products of *MATa* in the *a* cell (if any) are unknown. It may be required only to repress haploid functions in diploid cells.

- In α haploids, the α1 product turns on α-specific genes whose products are needed for α mating type. The α2 product represses the genes responsible for producing *a* mating type, by binding to an operator sequence located upstream of target genes.

- In diploids, the *a1* and α2 products cooperate to repress haploid-specific genes. They combine to recognize an operator sequence different from the target for α2 alone.

The abilities of the α2, *a1*, and α1 proteins to regulate transcription rely upon some interesting protein-protein interactions between themselves and with other protein(s). The pattern of gene control in *a* cells, α cells, and diploids, is summarized in **Figure 18.10**.

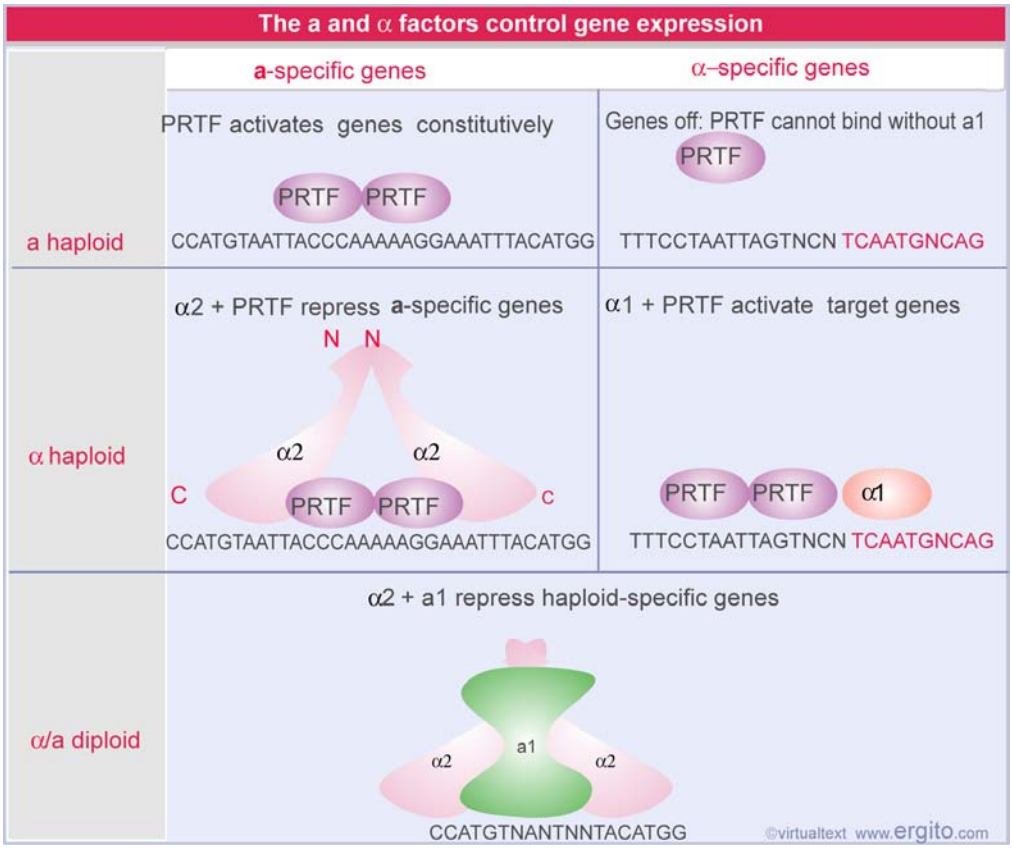

Figure 18.10 Combinations of PRTF, a1, α 1 and α 2 activate or repress specific groups of genes to correspond with the mating type of the cell.

A protein called PRTF (which is not specific for mating type) is involved in many of these interactions. PRTF binds to a short consensus sequence called the P box. The role of PRTF in gene regulation may be quite extensive, because P boxes are found in a variety of locations. In some of these sites, the P box is required for activation of the gene; but at other loci, PRTF is needed for repression. Its effects may therefore depend on the other proteins that bind at sites adjacent to the P box.

Genes that are *a*-specific may be activated by PRTF alone. This is adequate to ensure their expression in an *a* haploid.

The *a*-specific genes are repressed in an α haploid by the combined action of the α2 protein and PRTF. The α2 protein contains two domains. The C-terminal domain binds to short palindromic elements at the ends of an operator consensus sequence of 32 bp. However, binding of this fragment to DNA does not cause repression. The N-terminal domain is needed for repression and is responsible for making contacts with PRTF. The binding site for PRTF is a P box in the center of the operator. In fact, α and PRTF bind to the operator cooperatively.

Expression of α-specific genes requires the α1 activator. This is another small protein, 175 amino acids long. *Cis*-acting sequences that confer α-specific transcription are 26 bp long, and can be divided into two parts. The first 16 bp form the P box, where PRTF binds; the adjacent 10 bp sequence forms the binding site for

α1. The α1 factor binds only when PRTF is present to bind to the P box. Neither protein alone can bind to its target box, but together they can bind to DNA, presumably as a result of protein-protein interactions.

The α-specific genes are turned off by default in *a* haploids, because in the absence of α1 protein, PRTF is unable to bind to activate them.

The α2 protein can also cooperate with the α1 protein. The combination of these proteins recognizes a different operator. The operator shares the outlying palindromic sequences with the sequence recognized by α2 alone, but is shorter because the sequence between them is different. The α1/ α2 combination represses genes with this motif in diploid cells.

The major point to be made from these results is that the phenotype of each type of cell (*a* or α haploid or *a*/ α diploid) is determined by the combination of *a* and α proteins that are expressed. One aspect is the distinction between the haploid and diploid phenotypes; another is the distinction between *a* and α haploid phenotypes. The latter extends to expression of genes corresponding to the appropriate mating type and to the determination of the direction of switching of mating type (see **Figure 18.7**). *MATa* cells activate a recombination enhancer on the left arm of chromosome III, which increases recombination over a 40 kb region that includes *HML*. *MAT* α cells inactivate the left end of chromosome III (for review see 173).

Reviews

173. Nasmyth, K. and Shore, D. (1987). *Transcriptional regulation in the yeast life cycle.* Science 237, 1162-1170.

REARRANGEMENT OF DNA

4.18.7 Silent cassettes at *HML* and *HMR* are repressed

Key Terms

A **silencer** is a short sequence of DNA that can inactivate expression of a gene in its vicinity.

Key Concepts

- *HML* and *HMR* are repressed by silencer elements.
- Loci required to maintain silencing include *SIR1-4*, *RAP1*, and genes for histone H4.
- Binding of ORC (origin recognition complex) at the silencers is necessary for inactivation.

The transcription map in **Figure 18.8** reveals an intriguing feature. Transcription of either *MATa* or *MAT* α initiates within the *Y* region. Only the *MAT* locus is expressed; yet the same *Y* region is present in the corresponding nontranscribed cassette (*HML* or *HMR*). This implies that regulation of expression is not accomplished by direct recognition of some site overlapping with the promoter. *A site outside the cassettes must distinguish HML and HMR from MAT.*

Deletion analysis shows that sites on either side of both *HML* and *HMR* are needed to repress their expression. They are called **silencers**. The sites on the left of each cassette are called the E silencers, and the sites on the right side are called the I silencers. These control sites can function at a distance (up to 2.5 kb away from a promoter) and in either orientation. They behave like negative enhancers (enhancers are elements distant from the promoter that activate transcription; see *Molecular Biology 5.21.15 Enhancers contain bidirectional elements that assist initiation*; for review see 175).

Can we find the basis for the control of cassette activity by identifying genes that are responsible for keeping the cassettes silent? We would expect the products of these genes to act on the silencers. A convenient assay for mutation in such genes is provided by the fact that, when a mutation allows the usually silent cassettes at *HML* and *HMR* to be expressed, both *a* and α functions are produced, so the cells behave like *MATa/MAT* α diploids.

Mutations in several loci abolish silencing and lead to expression of *HML* and *HMR*. The first to be discovered were the four *SIR* loci (silent information regulators). All four wild-type *SIR* loci are needed to maintain *HML* and *HMR* in the repressed state; mutation in any one of these loci to give a sir^- allele has two effects. Both *HML* and *HMR* can be transcribed. And both the silent cassettes become targets for replacement by switching. So the same regulatory event is involved in repressing a

silent cassette and in preventing it from being a recipient for replacement by another cassette.

Other loci required for silencing include *RAP1* (which is also required to maintain telomeric heterochromatin in its inert state) and the genes coding for histone H4. Deletions of the N-terminus of histone H4 or individual point mutations activate the silent cassettes. The effects of these mutations can be overcome either by introducing new mutations in *SIR3* or by over-expressing *SIR1*, which suggests that there is a specific interaction between H4 and the SIR proteins.

The general model suggested by these results is that the SIR proteins act on chromatin structure to prevent expression of the genes. Because mutations in the SIR proteins have the same effects on genes that have been inactivated by the proximity of telomeric heterochromatin, it seems likely that SIR proteins are involved generally in interacting with histones to form heterochromatic (inert) structures (see *Molecular Biology 5.23.15 Heterochromatin depends on interactions with histones*).

There is an interesting connection between repression at the silencers and DNA replication. Each silencer contains an *ARS* sequence (an origin of replication). The *ARS* is bound by the ORC (the origin recognition complex) that is involved in initiating replication. Mutations in *ORC* genes prevent silencing, indicating that the binding of ORC protein at the silencer is required for silencing.

There are two separate types of connection between silencing and the replication apparatus:

- The presence of Sir1 is necessary.
- And replication is required.

If a Sir1 protein is localized at the silencer (by linkage to another protein that is bound there), the binding of ORC is no longer necessary. This means that the role of ORC is solely to bring in Sir1; it is not required to initiate replication. As illustrated in **Figure 18.11**, the role of ORC could therefore be to provide an initiating center from which the silencing effect can spread. ORC provides the structure to which Sir1p binds, and Sir1p then recruits the other SIR proteins. This is different from its role in replication.

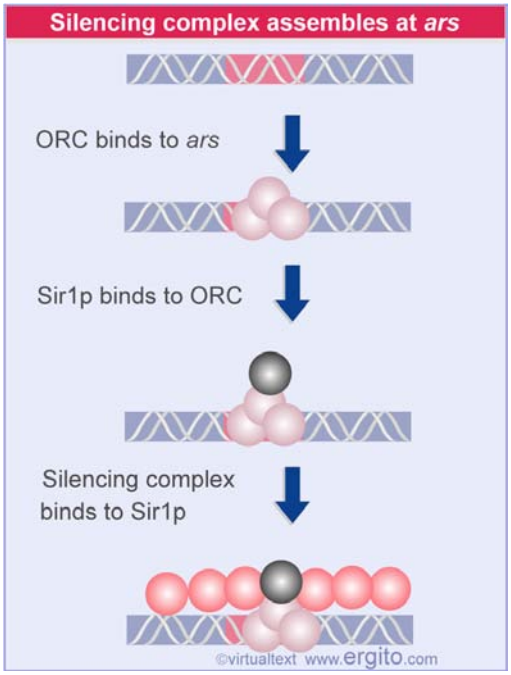

Figure 18.11 Silencing at HMR depends on recruiting Sir1p.

However, passage through S phase is necessary for silencing to be established. This does not require initiation to occur at the *ARS* in the silencer. The effect could depend on the passage of a replication fork through the silencer, perhaps in order to allow the chromatin structure to be changed.

Reviews

175. Laurenson, P. and Rine, J. (1992). *Silencers, silencing, and heritable transcriptional states.* Microbiol. Rev. 56, 543-560.

REARRANGEMENT OF DNA

4.18.8 Unidirectional transposition is initiated by the recipient *MAT* locus

Key Concepts

- Mating type switching is initiated by a double strand break made at the *MAT* locus by the HO endonuclease.

A switch in mating type is accomplished by a gene conversion in which the recipient site (*MAT*) acquires the sequence of the donor type (*HML* or *HMR*). Sites needed for transposition have been identified by mutations at *MAT* that prevent switching. The unidirectional nature of the process is indicated by lack of mutations in *HML* or *HMR*.

The mutations identify a site at the right boundary of *Y* at *MAT* that is crucial for the switching event. The nature of the boundary is shown by analyzing the locations of these point mutations relative to the site of switching (this is done by examining the results of rare switches that occur in spite of the mutation). Some mutations lie within the region that is replaced (and thus disappear from *MAT* after a switch), while others lie just outside the replaced region (and therefore continue to impede switching). So sequences both within and outside the replaced region are needed for the switching event.

Switching is initiated by a double-strand break close to the *Y-Z* boundary that coincides with a site that is sensitive to attack by DNAase. (This is a common feature of chromosomal sites that are involved in initiating transcription or recombination.) It is recognized by an endonuclease coded by the *HO* locus. The HO endonuclease makes a staggered double-strand break just to the right of the *Y* boundary. Cleavage generates the single-stranded ends of 4 bases drawn in **Figure 18.12**. The nuclease does not attack mutant *MAT* loci that cannot switch. Deletion analysis shows that most or all of the sequence of 24 bp surrounding the *Y* junction is required for cleavage *in vitro*. The recognition site is relatively large for a nuclease, and it occurs only at the three mating-type cassettes (583).

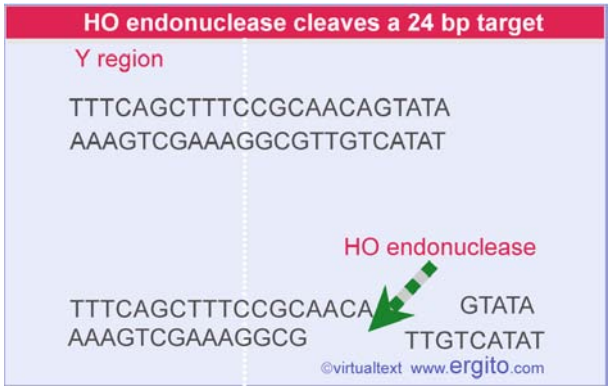

Figure 18.12 HO endonuclease cleaves *MAT* just to the right of the Y region, generating sticky ends with a base overhang.

Only the *MAT* locus and not the *HML* or *HMR* loci are targets for the endonuclease. It seems plausible that the same mechanisms that keep the silent cassettes from being transcribed also keep them inaccessible to the HO endonuclease. This inaccessibility ensures that switching is unidirectional.

The reaction triggered by the cleavage is illustrated schematically in **Figure 18.13** in terms of the general reaction between donor and recipient regions. In terms of the interactions of individual strands of DNA, it follows the scheme for recombination via a double-strand break drawn in **Figure 15.8**; and the stages following the initial cut require the enzymes involved in general recombination. Mutations in some of these genes prevent switching.

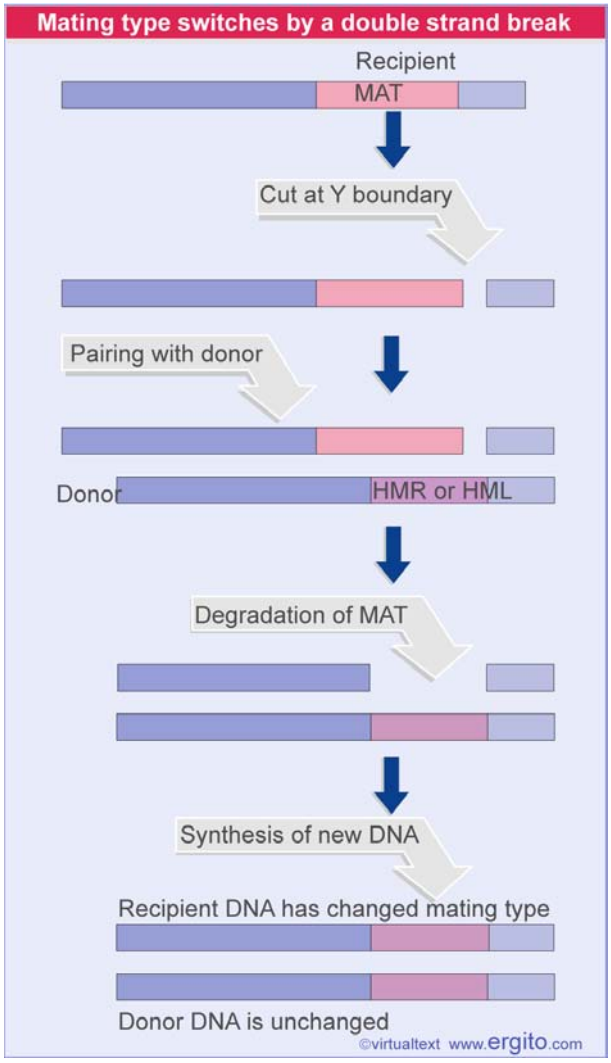

Figure 18.13 Cassette substitution is initiated by a double-strand break in the recipient (*MAT*) locus, and may involve pairing on either side of the Y region with the donor (*HMR* or *HML*) locus.

Suppose that the free end of *MAT* invades either the *HML* or *HMR* locus and pairs with the region of homology on the right side. The *Y* region of *MAT* is degraded until a region with homology on the left side is exposed. At this point, *MAT* is paired with *HML* or *HMR* at both the left side and the right side. The *Y* region of *HML* or *HMR* is copied to replace the region lost from *MAT* (which might extend beyond the limits of *Y* itself). The paired loci separate. (The order of events could be different.)

Like the double-strand break model for recombination, the process is initiated by *MAT*, the locus that is to be replaced. In this sense, the description of *HML* and *HMR* as donor loci refers to their ultimate role, but not to the mechanism of the process. Like replicative transposition, the donor site is unaffected, but a change in sequence occurs at the recipient; unlike transposition, the recipient locus suffers a substitution rather than addition of material.

References

583. Strathern, J. N., Klar, A. J., Hicks, J. B., Abraham, J. A., Ivy, J. M., Nasmyth, K. A., and McGill, C. (1982). *Homothallic switching of yeast mating type cassettes is initiated by a double-stranded cut in the MAT locus.* Cell 31, 183-192.

REARRANGEMENT OF DNA

4.18.9 Regulation of HO expression controls switching

Key Concepts

- HO endonuclease is synthesized in haploid mother cells, so that a switching event causes both daughters to have the new mating type.

Production of the HO endonuclease is regulated at the level of gene transcription. There are three separate control systems:

- *HO* is under mating-type control. It is not synthesized in *MATa*/*MAT* α diploids. The reason could be that there is no need for switching when both *MAT* alleles are expressed anyway.
- *HO* is transcribed in mother cells but not in daughter cells.
- *HO* transcription also responds to the cell cycle. The gene is expressed only at the end of the G1 phase of a mother cell.

The timing of nuclease production explains the relationship between switching and cell lineage. **Figure 18.14** shows that switching is detected only in the products of a division; both daughter cells have the same mating type, switched from that of the parent. The reason is that the restriction of *HO* expression to G1 phase ensures that the mating type is switched before the *MAT* locus is replicated, with the result that both progeny have the new mating type.

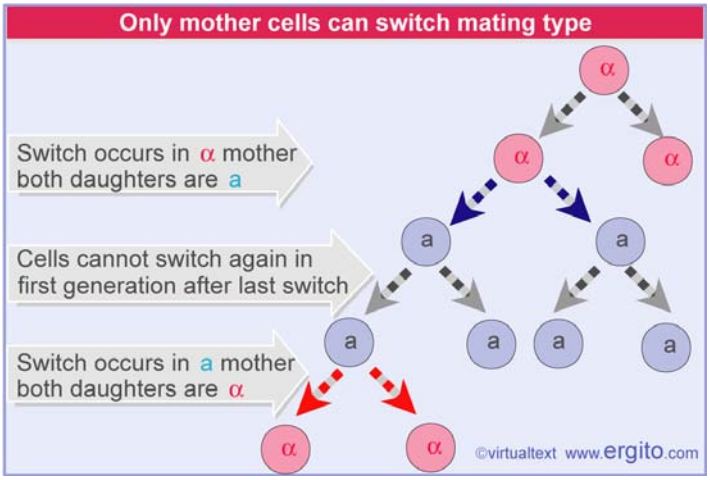

Figure 18.14 Switching occurs only in mother cells; both daughter cells have the new mating type. A daughter cell must pass through an entire cycle before it becomes a mother cell that is able to switch again.

cis-acting sites that control *HO* transcription reside in the 1500 bp upstream of the gene. The general pattern of control is that repression at any one of many sites, responding to several regulatory circuits, may prevent transcription of *HO*. **Figure 18.15** summarizes the types of sites that are involved.

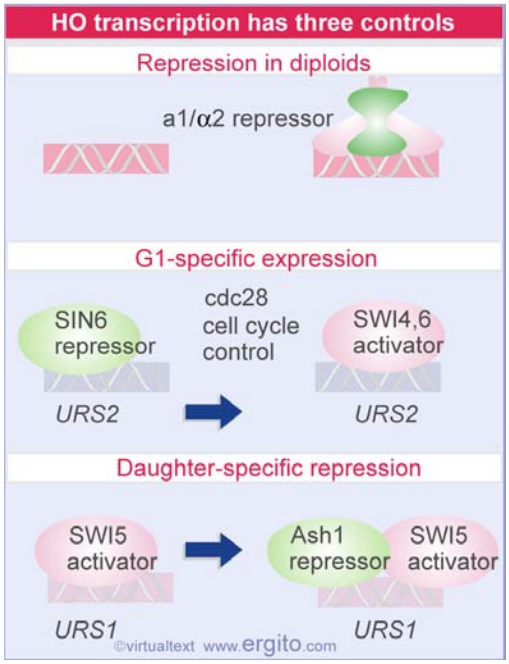

Figure 18.15 Three regulator systems act on transcription of the *HO* gene. Transcription occurs only when all repression is lifted.

Mating type control resembles that of other haploid-specific genes. Transcription is prevented (in diploids) by the *a1*/ α2 repressor. There are 10 binding sites for the repressor in the upstream region. These sites vary in their conformity to the consensus sequence; we do not know which and how many of them are required for haploid-specific repression.

The control of *HO* transcription involves interplay between a series of activating and repressing events. The genes *SWI1-5* are required for *HO* transcription. They function by preventing products of the genes *SIN1-6* from repressing *HO*. The *SWI* genes were discovered first, as mutants unable to switch; then the *SIN* genes were discovered for their ability to release the blocks caused by particular *SWI* mutations. *SWI-SIN* interactions are involved in both cell-cycle control and the restriction of expression to mother cells.

Some of the *SWI* and *SIN* genes are not specifically concerned with mating type, but are global regulators of transcription, whose functions are needed for expression of many loci. They include the activator complex *SWI1,2,3* and the loci *SIN1-4* that code for chromosomal proteins. Their role in mating type expression is incidental. The "real" regulator is therefore the SWI protein that counteracts the general repression system specifically at the *HO* locus.

Cell-cycle control is conferred by 9 copies of an octanucleotide sequence called

URS2. A copy of the consensus sequence can confer cell-cycle control on a gene to which it is attached. A gene linked to this sequence is repressed except during a transient period toward the end of G1 phase. SWI4 and SWI6 are the activators that release repression at *URS2*. Their activity depends on the function of the cell-cycle regulator CDC28, which executes the decision that commits the cell to divide (see *Molecular Biology 6.29 Cell cycle and growth regulation*).

The target for restricting expression to alternate generations is the activator SWI5 (which antagonizes a general repression system exercised by SIN3,4). In mutants that lack these functions, *HO* is transcribed equally well in mother and daughter cells. This system acts on *URS1* elements in the far upstream region.

SWI5 is not itself the regulator of mother-cell specificity, but is antagonized by Ash1p, a repressor that accumulates preferentially in daughter cells at the end of anaphase. Mutations in *ASH1* allow daughter cells to switch mating type. The localization of Ash1p is determined by the transport of its mRNA from the mother cell along actin filaments into the daughter bud (*Molecular Biology 2.5.16 mRNA can be specifically localized*). Its presence prevents SWI5 from activating the *HO* gene. It works by binding to many copies of a consensus sequence that are distributed throughout the regulatory regions *URS1* and *URS2* (3154). When the daughter cell grows to become a mother cell, the concentration of Ash1p is diluted, and it becomes possible to express the *HO* gene again (585).

References

585. Bobola, N., Jansen, R. P., Shin, T. H., and Nasmyth, K. (1996). *Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells.* Cell 84, 699-709.

3154. Maxon, M. E. and Herskowitz, I. (2001). *Ash1p is a site-specific DNA-binding protein that actively represses transcription.* Proc. Natl. Acad. Sci. USA 98, 1495-1500.

REARRANGEMENT OF DNA

4.18.10 Trypanosomes switch the VSG frequently during infection

Key Terms

The **variable surface glycoprotein (VSG)** is the protein on the surface of a trypanosome that changes during an infection so as to prevent the infected host from mounting an immune reaction to it.

Antigenic variation describes the ability of a trypanosome to change its surface protein, so that the host is challenged with a different antigen.

Key Concepts

- The trypanosome life cycle alternates between tsetse fly and mammal.
- The form of the parasite that is transmitted to the mammal has a coat of a VSG (variable surface glycoprotein).
- The VSG is replaced every 1-2 weeks.

Sleeping sickness in man (and a related disease in cows) is caused by infection with African trypanosomes. The unicellular parasite follows the life cycle illustrated in **Figure 18.16**, in which it alternates between tsetse fly and mammal. The trypanosome may be transferred either to or from the fly when it bites a mammal.

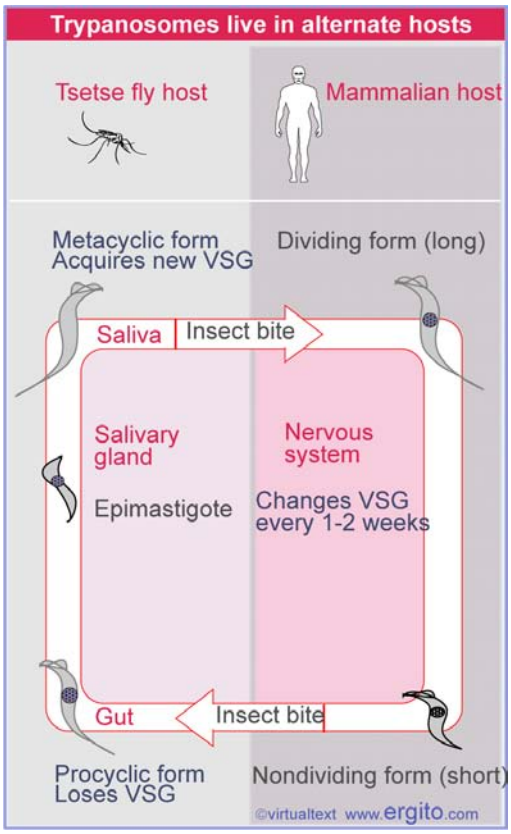

Figure 18.16 A trypanosome passes through several morphological forms when its life cycle alternates between a tsetse fly and mammalian host.

During its life cycle, the parasite undergoes several morphological and biochemical changes. The most significant biochemical change is in the **variable surface glycoprotein** (**VSG**), the major component of the surface coat. The coat covers the plasma membrane and consists of a monolayer of 5-10 × 10^6 molecules of a single VSG, which is the only antigenic structure exposed on the surface. A trypanosome expresses only one VSG at any time, and its ability to change the VSG is responsible for its survival through the fly-mammal infective cycle.

Consider the cycle as starting when a fly gains a trypanosome by biting an infected mammal. The trypanosome enters the gut of the fly in the "procyclic form," and loses its VSG. After about three weeks, its progeny differentiate into the "metacyclic form," which re-acquires a VSG coat. This form is transmitted to the mammalian bloodstream during a bite by the fly. The trypanosome multiplies in the mammalian bloodstream. Its progeny continue to express the metacyclic VSG for about a week. Then a new VSG is synthesized, and further transitions occur every 1-2 weeks.

Each of the successive VSG species is immunologically distinct. As a result, the antigen presented to the mammalian immune system is constantly changing. The process of transition is called **antigenic variation**. The immune response of the organism always lags behind the change in surface antigen, so that the trypanosome evades immune surveillance, and thereby perpetuates itself indefinitely. Each transition of the VSG is accompanied by a new wave of parasitemia, with symptoms

of fever, rash, etc.; the parasites eventually invade the central nervous system, after which the mammalian host becomes progressively more lethargic and eventually comatose.

Trypanosomes vary in their host range. The best investigated species is a variety of *Trypanosoma brucei* that grows well on laboratory animals (although not on man). Laboratory strains of *T. brucei* switch VSGs spontaneously at a rate of 10^{-4} - 10^{-6} per division. Switching occurs independently of the host immune system. In effect, new variants then are selected by the host, because it mounts a response against the old VSG, but fails to recognize and act against the new VSG (for review see 177; 178; 2515).

A general view of VSG structure is depicted in **Figure 18.17**. A nascent VSG is ~500 amino acids long; it has an N-terminal signal sequence, followed by a long variable region that provides the unique antigenic determinant, and a C-terminal homology region ending in a short hydrophobic tail. The nascent VSG is processed at both ends to give the mature form. The signal sequence is cleaved during secretion. The hydrophobic tail is removed before the VSG reaches the outside surface. The new C-terminus is covalently attached to the trypanosome membrane; three types of homology region are distinguished according to the C-terminal amino acid.

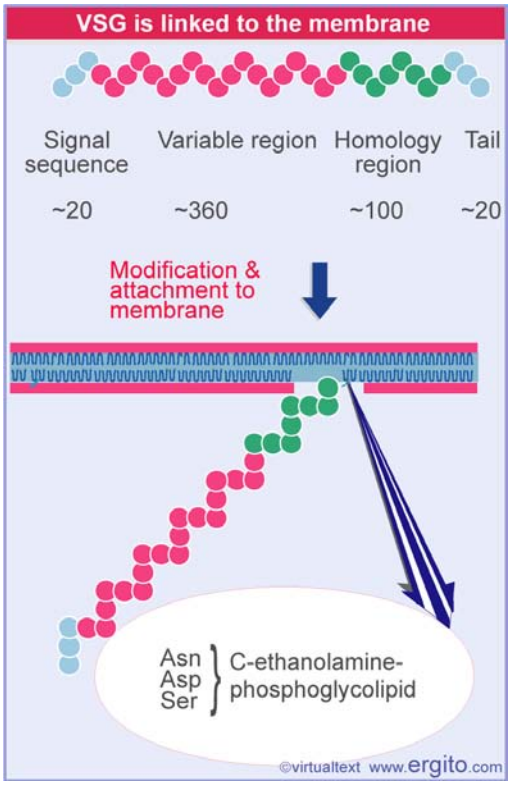

Figure 18.17 The C-terminus of VSG is cleaved and covalently linked to the membrane through a glycolipid.

The VSG is attached to the membrane via a phosphoglycolipid. As a result, VSG can

be released from the membrane by an enzyme that removes fatty acid. This reaction (which is used in purifying the VSG) may be important *in vivo* in allowing one VSG to be replaced by another on the surface of the trypanosome.

Reviews

177. Boothroyd, J. C. (1985). *Antigenic variation in African trypanosomes.* Annu. Rev. Immunol. 39, 475-502.

178. Donelson, J. E. and Rice-Ficht, A. C. (1985). *Molecular biology of trypanosome antigenic variation.* Microbiol. Rev. 49, 107-125.

2515. Barry, J. D., McCulloch, R., and Barry, R. (2001). *Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite.* Adv. Parasitol. 49, 1-70.

REARRANGEMENT OF DNA

4.18.11 New VSG sequences are generated by gene switching

Key Terms

Each VSG (variable surface glycoprotein) of a trypanosome is coded by a **basic copy gene**.

The **expression-linked copy in (ELC)** a trypanosome genome is the one copy of a VSG gene that is expressed.

An **expression site** in a trypanosome genome is a locus near a telomere that can express the VSG gene that is located there.

Key Concepts

- The trypanosome has ~1000 basic copy VSG genes.
- Only a single expression-linked copy located near a telomere is expressed at any given time.
- A basic copy gene is activated by having its sequence copied into the expression site.
- There are only a few potential expression sites.

How many varieties of VSG can be expressed by any one trypanosome? It is not clear that any limit is encountered before death of the host. A single trypanosome can make at least 100 VSGs sufficiently different in sequence that antibodies against any one do not react against the others.

VSG variation is coded in the trypanosome genome. Every individual trypanosome carries the entire VSG repertoire of its strain. Diversity therefore depends on changing expression from one preexisting gene to another.

The trypanosome genome has an unusual organization, consisting of a large number of segregating units. In addition to an unknown number of chromosomes, it contains ~100 "minichromosomes," each containing ~50-150 kb of DNA. Hybridization experiments identify ~1000 VSG genes, scattered among all size classes of chromosomal material.

Each VSG is coded by a **basic copy gene**. These genes can be divided into two classes according to their chromosomal location:

- Telomeric genes lie within 5-15 kb of a telomere. There could be >200 of these genes if every telomere has one.
- Internal genes reside within chromosomes (more formally, they lie >50 kb from

a telomere).

As might be expected of a large family of genes, individual basic copies show varying degrees of relationship, presumably reflecting their origin by duplication and variation. Genes that are closely related, and which provoke the same antigenic response, are called isogenes.

How is a single VSG gene selected for expression? Only one VSG gene is transcribed in a trypanosome at a given time. The copy of the gene that is active is called the **expression-linked copy** (**ELC**). It is said to be located at an **expression site**. An expression site has a characteristic property: it is located near a telomere.

These features immediately suggest that the route followed to select a gene for expression depends on whether the basic copy is itself telomeric or internal. The two types of event that can create an ELC are summarized in **Figure 18.18**:

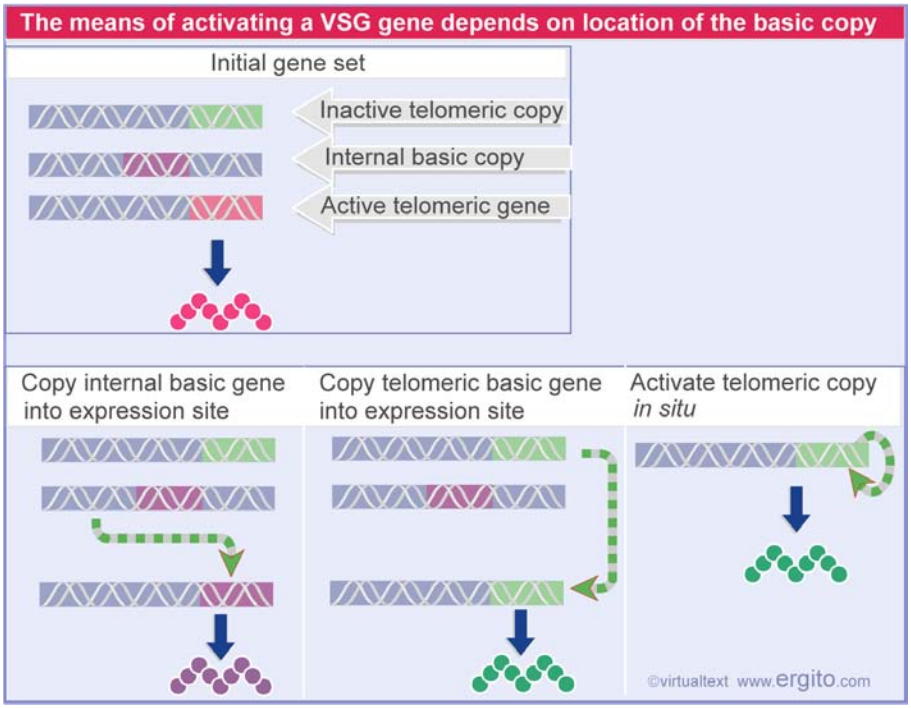

Figure 18.18 VSG genes may be created by duplicative transfer from an internal or telomeric basic copy into an expression site, or by activating a telomeric copy that is already present at a potential expression site.

- The expression site remains the same, but the ELC is changed. Duplication transfers the sequence of a basic copy to replace the sequence currently occupying the expression site. Either internal or telomeric copies may be activated directly by duplication into the expression site. The substitution of one cassette for another does not interfere with the activity of the site.

- The expression site is changed. Activation *in situ* is available only to a sequence already present at a telomere. When a telomeric site is activated *in situ*, the

previous expression site must cease to be active and the new site now becomes the expression site.

Internal basic copies probably can be copied into non-expressed telomeric locations as well as into expression sites. So an internal gene could be activated by a two-stage process, in which first it is transposed to a non-expressed telomere, and then this site is activated.

We can follow the fate of genes involved in activation by restriction mapping. A probe representing an expressed sequence can be derived from the mRNA. Then we can determine the status of genes corresponding to the probe. We see different results for internal and telomeric basic copy genes:

- Activation of an internal gene requires generation of new sequences. **Figure 18.19** shows that when an internal gene is activated, a new fragment is found. The original basic-copy gene remains unaltered; the new fragment is generated by the duplication of the gene into a new context (where the sites recognized by the restriction enzyme are in the surrounding sequences and therefore generate a distinct fragment). The new fragment identifies an ELC, located close to a telomere. The ELC appears when the gene is expressed and disappears when the gene is switched off. Duplication into the ELC is the only pathway by which an internal basic copy can be generated.

- Activation of a telomeric gene can occur *in situ*. **Figure 18.20** shows that when a telomeric gene is activated, the gene number need not change. The structure of the gene may be essentially unaffected as detected by restriction mapping. The size of the fragment containing the gene may vary slightly, because the length of the telomere is constantly changing. Telomeric basic copies can also be activated by the same duplication pathway as internal copies; in this case, the basic copy remains at its telomere, while an expression-linked copy appears at another telomere (generating a new fragment as illustrated for internal basic copies in **Figure 18.19**).

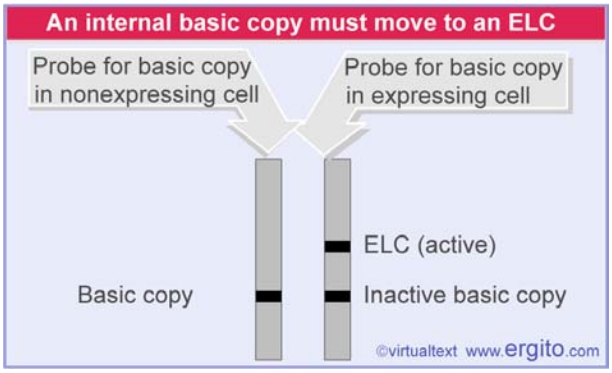

Figure 18.19 Internal basic copies can be activated only by generating a duplication of the gene at an expression-linked site

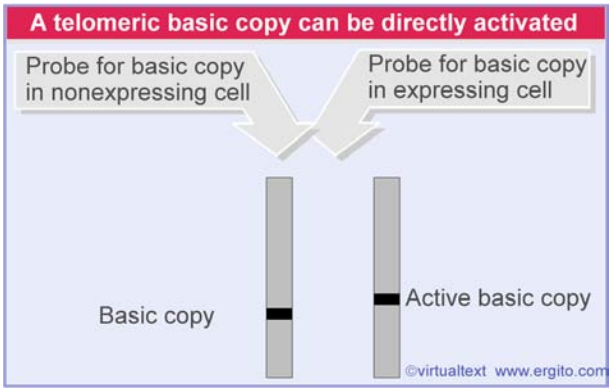

Figure 18.20 Telomeric basic copies can be activated *in situ*; the size of the restriction fragment may change (slightly) when the telomere is extended.

Formation of the ELC occurs by a gene conversion process that requires genetic recombination enzymes – for example, it is greatly reduced by mutation in *RAD51* (3156). Like the switch in yeast mating type, it represents the replacement of a "cassette" at the active (telomeric) locus by a stored cassette. The VSG system is more versatile in the sense that there are many potential donor cassettes (and also more than a single potential recipient site).

Almost all switches in VSG type involve replacement of the ELC by a pre-existing silent copy. Some exceptional cases have been found, however, in which the sequence of the ELC does not match any of the repertoire of silent copies in the genome. A new sequence may be created by a series of gene conversions in which short stretches of different silent copies are connected. This resembles the mechanism for generating diversity in chicken λ immunoglobulins (see *Molecular Biology 5.25.16 Avian immunoglobulins are assembled from pseudogenes*). Although rare, such occurrences extend VSG diversity.

There are ~20 potential expression sites, which means that only a subset of telomeres can function in this capacity. All the expression sites appear similar in sequence and organization. Switching from one expression site to another occurs at a low frequency. This is not a principal means of changing coat expression, but has the effect of switching expression of other genes located within the expression site. Among these genes are two coding for the subunits of a (heterodimeric) transferrin receptor. Changing the transferrin receptor that is expressed by the trypanosome has a strong effect on its affinity for the host transferrin protein.

References

3156. McCulloch, R. and Barry, J. D. (1999). *A role for RAD51 and homologous recombination in Trypanosoma brucei antigenic variation.* Genes Dev. 13, 2875-2888.

REARRANGEMENT OF DNA

4.18.12 VSG genes have an unusual structure

Key Concepts

- The coding region of the VSG gene is flanked by barren regions of repetitive DNA.
- The VSG sequence is transferred into the ELC between the promoter and terminator.
- The 5′ end of VSG mRNA is added by a *trans*-splicing reaction to the 5′ end that is generated by transcription.
- Activation of an expression site involves a change in the region upstream of the promoter.
- The expression site may be contained in a special extranucleolar body where the VSG is transcribed by RNA polymerase I.

The structure of the VSG gene at the ELC is unusual, as illustrated in **Figure 18.21**. The length of DNA transferred into the ELC is 2500-3500 bp, somewhat longer than the VSG-coding region of 1500 bp. Most of the additional length is upstream of the gene. The crossover points at which the duplicated sequence joins the ELC do not appear to be precisely determined (for review see 179).

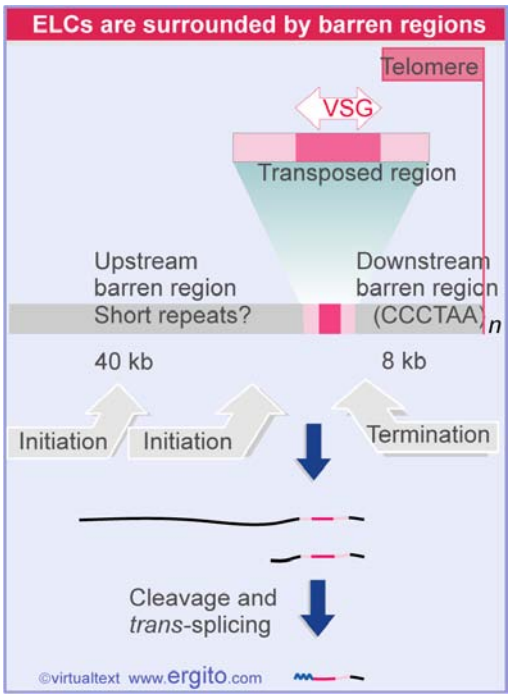

Figure 18.21 The expression-linked copy of a VSG gene contains barren regions on either side of the transposed region, which extends from ~1000 bp upstream of the VSG coding region to a site near the 3′ terminus of the mRNA.

Analysis of events at the 5′ end of the VSG mRNA is complicated by the fact that the mature RNA starts with a 35 base sequence coded elsewhere, and added in *trans* to the newly synthesized 5′ end (see *Molecular Biology 5.24.13 trans-splicing reactions use small RNAs*). The signals for initiating and terminating transcription (and sometimes also the end of the coding region itself) are provided by the sequences flanking the transposed region. In fact, transcription may be initiated several kb upstream of the VSG gene itself. Promoters have been mapped at 4 kb and ~60 kb upstream of the VSG sequence. Use of the more distant promoter generates a transcript that contains other genes as well as the active VSG. The VSG sequence (and other gene sequences) must be released by cleavage from the transcript, after which the 35 base spliced leader is added to the 5′ end.

The RNA polymerase that transcribes the expression locus is not the usual RNA polymerase II, but is RNA polymerase I (the enzyme that usually transcribes rRNA.) The ELC is sequestered in a discrete nuclear body, called the expression site body (ESB) (2517). The ESB takes the form of an extranucleolar body containing RNA polymerase I, and is found only in the bloodstream form. This may explain why only one of the potential 20 expression copies is in fact expressed in a given trypanosome. If the ESB is necessary for expression and can only accommodate a single copy, then by default it will prevent expression of all the other copies.

On either side of the transposed region are extensive regions that are not cut by restriction enzymes. These "barren regions" consist of repetitive DNA; they extend some 8 kb downstream and for up to 40 kb upstream of the ELC. Going downstream, the barren region consists largely of repeats of the sequence CCCTAA, and extends

to the telomere. Proceeding upstream, it may also consist of repetitive sequences, but their nature is not yet clear. The existence of the barren regions, however, has been an impediment to characterizing ELC genes by cloning.

The order in which VSG genes are expressed during an infection is erratic, but not completely random. This may be an important feature in survival of the trypanosome. If VSG genes were used in a predetermined order, a host could knock out the infection by mounting a reaction against one of the early elements. The need for unpredictability in the production of VSGs may be responsible for the evolution of a system with many donor sequences and multiple recipients.

Antigenic variation is not a unique phenomenon of trypanosomes. The bacterium *Borrelia hermsii* causes relapsing fever in man and analogous diseases in other mammals. The name of the disease reflects its erratic course: periods of illness are spaced by periods of relief. When the fevers occur, spirochetes are found in the blood; they disappear during periods of relief, as the host responds with specific antibodies.

Like the trypanosomes, *Borrelia* survives by altering a surface protein, called the variable major protein (VMP). Changes in the VMP are associated with rearrangements in the genome. The active VMP is located near the telomere of a linear plasmid. We do not yet know the extent of the coded variants or the mechanisms used to alter their expression. It is intriguing, however, that the eukaryote *Trypanosoma* and the prokaryote *Borrelia* should both rely upon antigenic variation as a means for evading immune surveillance.

Last updated on 6-20-2002

Reviews

179. Borst, P. (1986). *Discontinuous transcription and antigenic variation in trypanosomes.* Annu. Rev. Biochem. 55, 701-732.

References

2517. Navarro, M. and Gull, K. (2001). *A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei.* Nature 414, 759-763.

REARRANGEMENT OF DNA

4.18.13 The bacterial Ti plasmid causes crown gall disease in plants

Key Terms

Crown gall disease is a tumor that can be induced in many plants by infection with the bacterium *Agrobacterium tumefaciens*.

The **Ti plasmid** is an episome of the bacterium *Agrobacterium tumefaciens* that carries the genes responsible for the induction of crown gall disease in infected plants.

An **opine** is a derivative of arginine that is synthesized by plant cells infected with crown gall disease.

Nopaline plasmids are Ti plasmids of *Agrobacterium tumefaciens* that carry genes for synthesizing the opine, nopaline. They retain the ability to differentiate into early embryonic structures.

A **teratoma** is a growth in which many differentiated cell types - including skin, teeth, bone and others - grow in disorganized manner after an early embryo is transplanted into one of the tissues of an adult animal.

Octopine plasmids of *Agrobacterium tumefaciens* carry genes coding the synthesis of opines of the octopine type. The tumors are undifferentiated.

Agropine plasmids carry genes coding for the synthesis of opines of the agropine type. The tumors usually die early.

Ri plasmids are found in *Agrobacterium tumefaciens*. Like Ti plasmids, they carry genes that cause disease in infected plants. The disease may take the form of either hairy root disease or crown gall disease.

Key Concepts

- Infection with the bacterium *A. tumefaciens* can transform plant cells into tumors.
- The infectious agent is a plasmid carried by the bacterium.
- The plasmid also carries genes for synthesizing and metabolizing opines (arginine derivatives) that are used by the tumor cell.

Most events in which DNA is rearranged or amplified occur within a genome, but the interaction between bacteria and certain plants involves the transfer of DNA from the bacterial genome to the plant genome. **Crown gall disease**, shown in **Figure 18.22**, can be induced in most dicotyledonous plants by the soil bacterium *Agrobacterium tumefaciens*. The bacterium is a parasite that effects a genetic change in the eukaryotic host cell, with consequences for both parasite and host. It improves conditions for survival of the parasite. And it causes the plant cell to grow as a tumor.

Figure 18.22 An *Agrobacterium* carrying a Ti plasmid of the nopaline type induces a teratoma, in which differentiated structures develop. Photograph kindly provided by Jeff Schell.

Agrobacteria are required to induce tumor formation, but the tumor cells do not require the continued presence of bacteria. Like animal tumors, the plant cells have been transformed into a state in which new mechanisms govern growth and differentiation. Transformation is caused by the expression within the plant cell of genetic information transferred from the bacterium.

The tumor-inducing principle of *Agrobacterium* resides in the **Ti plasmid**, which is perpetuated as an independent replicon within the bacterium. The plasmid carries genes involved in various bacterial and plant cell activities, including those required to generate the transformed state, and a set of genes concerned with synthesis or utilization of **opines** (novel derivatives of arginine).

Ti plasmids (and thus the *Agrobacteria* in which they reside) can be divided into four groups, according to the types of opine that are made:

- **Nopaline** plasmids carry genes for synthesizing nopaline in tumors and for utilizing it in bacteria. Nopaline tumors can differentiate into shoots with abnormal structures. They have been called **teratomas** by analogy with certain mammalian tumors that retain the ability to differentiate into early embryonic structures.
- **Octopine** plasmids are similar to nopaline plasmids, but the relevant opine is different. However, octopine tumors are usually undifferentiated and do not form teratoma shoots.
- **Agropine** plasmids carry genes for agropine metabolism; the tumors do not differentiate, develop poorly, and die early.

- **Ri plasmids** can induce hairy root disease on some plants and crown gall on others. They have agropine type genes, and may have segments derived from both nopaline and octopine plasmids.

The types of genes carried by a Ti plasmid are summarized in **Figure 18.23**. Genes utilized in the bacterium code for plasmid replication and incompatibility, for transfer between bacteria, sensitivity to phages, and for synthesis of other compounds, some of which are toxic to other soil bacteria. Genes used in the plant cell code for transfer of DNA into the plant, for induction of the transformed state, and for shoot and root induction.

Ti genes function in bacteria and in plants

Locus	Function	Ti Plasmid
vir	DNA transfer into plant	all
shi	shoot induction	all
roi	root induction	all
nos	nopaline synthesis	nopaline
noc	nopaline catabolism	nopaline
ocs	octopine synthesis	octopine
occ	octopine catabolism	octopine
tra	bacterial transfer genes	all
Inc	incompatibility genes	all
oriV	origin for replication	all

Figure 18.23 Ti plasmids carry genes involved in both plant and bacterial functions.

The specificity of the opine genes depends on the type of plasmid. Genes needed for opine synthesis are linked to genes whose products catabolize the same opine; thus each strain of *Agrobacterium* causes crown gall tumor cells to synthesize opines that are useful for survival of the parasite. The opines can be used as the sole carbon and/or nitrogen source for the inducing *Agrobacterium* strain. The principle is that the transformed plant cell synthesizes those opines that the bacterium can use (for review see 182).

Reviews

182. Winans, S. C. (1993). *Two-way chemical signaling in Agrobacterium-plant interactions.* Microbiol. Rev. 56, 12-31.

REARRANGEMENT OF DNA

4.18.14 T-DNA carries genes required for infection

Key Terms

T-DNA is the segment of the Ti plasmid of *Agrobacterium tumefaciens* that is transferred to the plant cell nucleus during infection. It carries genes that transform the plant cell.

Key Concepts

- Part of the DNA of the Ti plasmid is transferred to the plant cell nucleus.
- The *vir* genes of the Ti plasmid are located outside the transferred region and are required for the transfer process.
- The *vir* genes are induced by phenolic compounds released by plants in response to wounding.
- The membrane protein VirA is autophosphorylated on histidine when it binds an inducer.
- VirA activates VirG by transferring the phosphate group to it.
- The VirA-VirG is one of several bacterial two component systems that use a phosphohistidine relay.

The interaction between *Agrobacterium* and a plant cell is illustrated in **Figure 18.24**. The bacterium does not enter the plant cell, but transfers part of the Ti plasmid to the plant nucleus. The transferred part of the Ti genome is called **T-DNA**. It becomes integrated into the plant genome, where it expresses the functions needed to synthesize opines and to transform the plant cell.

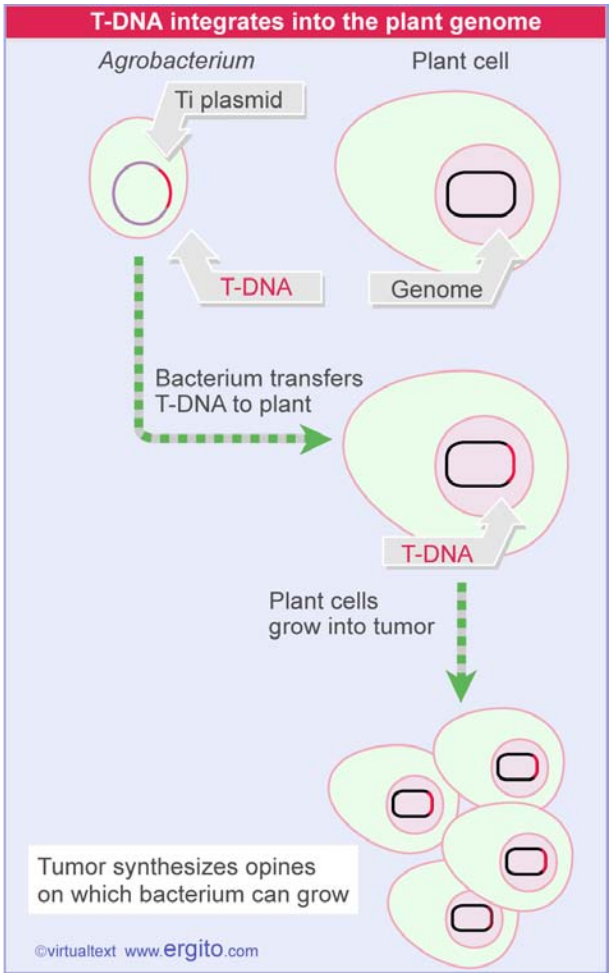

Figure 18.24 T-DNA is transferred from *Agrobacterium* carrying a Ti plasmid into a plant cell, where it becomes integrated into the nuclear genome and expresses functions that transform the host cell.

Transformation of plant cells requires three types of function carried in the *Agrobacterium*:

- Three loci on the *Agrobacterium* chromosome, *chvA*, *chvB*, *pscA*, are required for the initial stage of binding the bacterium to the plant cell. They are responsible for synthesizing a polysaccharide on the bacterial cell surface.
- The *vir* region carried by the Ti plasmid outside the T-DNA region is required to release and initiate transfer of the T-DNA.
- The T-DNA is required to transform the plant cell.

The organization of the major two types of Ti plasmid is illustrated in **Figure 18.25**. About 30% of the ~200 kb Ti genome is common to nopaline and octopine plasmids. The common regions include genes involved in all stages of the interaction between *Agrobacterium* and a plant host, but considerable rearrangement of the sequences has

occurred between the plasmids.

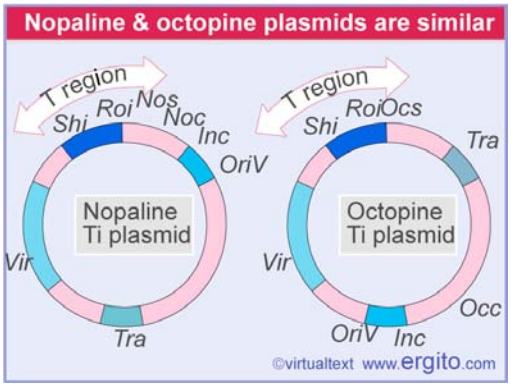

Figure 18.25 Nopaline and octopine Ti plasmids carry a variety of genes, including T-regions that have overlapping functions

The T-region occupies ~23 kb. Some 9 kb is the same in the two types of plasmid. The Ti plasmids carry genes for opine synthesis (*Nos* or *Ocs*) within the T-region; corresponding genes for opine catabolism (*Noc* or *Occ*) reside elsewhere on the plasmid. The plasmids code for similar, but not identical, morphogenetic functions, as seen in the induction of characteristic types of tumors.

Functions affecting oncogenicity – the ability to form tumors – are not confined to the T-region. Those genes located outside the T-region must be concerned with establishing the tumorigenic state, but their products are not needed to perpetuate it. They may be concerned with transfer of T-DNA into the plant nucleus or perhaps with subsidiary functions such as the balance of plant hormones in the infected tissue. Some of the mutations are host-specific, preventing tumor formation by some plant species, but not by others.

The virulence genes code for the functions required for the transfer process. Six loci *virA-G* reside in a 40 kb region outside the T-DNA. Their organization is summarized in **Figure 18.26**. Each locus is transcribed as an individual unit; some contain more than one open reading frame.

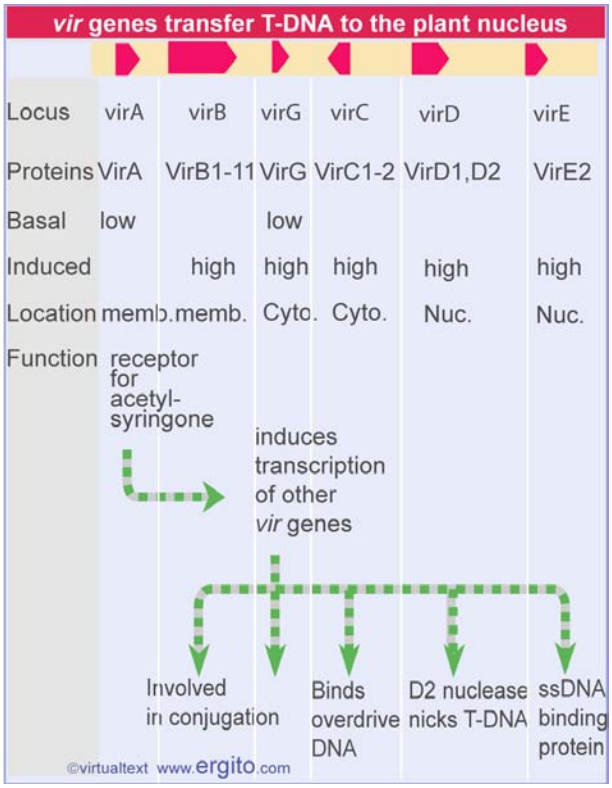

Figure 18.26 The *vir* region of the Ti plasmid has six loci that are responsible for transferring T-DNA to an infected plant.

We may divide the transforming process into (at least) two stages:

- Agrobacterium contacts a plant cell, and the *vir* genes are induced.
- *vir* gene products cause T-DNA to be transferred to the plant cell nucleus, where it is integrated into the genome.

The *vir* genes fall into two groups, corresponding to these stages. Genes *virA* and *virG* are regulators that respond to a change in the plant by inducing the other genes. So mutants in *virA* and *virG* are avirulent and cannot express the remaining *vir* genes. Genes *virB,C,D,E* code for proteins involved in the transfer of DNA. Mutants in *virB* and *virD* are avirulent in all plants, but the effects of mutations in *virC* and *virE* vary with the type of host plant.

virA and *virG* are expressed constitutively (at a rather low level). The signal to which they respond is provided by phenolic compounds generated by plants as a response to wounding. **Figure 18.27** presents an example. *N. tabacum* (tobacco) generates the molecules acetosyringone and α-hydroxyacetosyringone. Exposure to these compounds activates *virA*, which acts on *virG*, which in turn induces the expression de novo of *virB,C,D,E*. This reaction explains why *Agrobacterium* infection succeeds only on wounded plants.

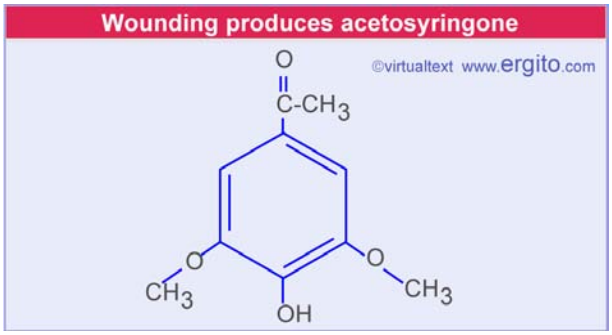

Figure 18.27 Acetosyringone (4-acetyl-2,6-dimethoxyphenol) is produced by *N. tabacum* upon wounding, and induces transfer of T-DNA from *Agrobacterium*.

VirA and VirG are an example of a classic type of bacterial system in which stimulation of a sensor protein causes autophosphorylation and transfer of the phosphate to the second protein (see *Molecular Biology Supplement 32.13 Two-component signal transduction*). The relationship is illustrated in **Figure 18.28**. The VirA-VirG system resembles the EnvZ-OmpR system that responds to osmolarity. The sequence of *virA* is related to *envZ*; and the sequences of *virG* and *ompR* are closely related, suggesting that the effector proteins function in a similar manner.

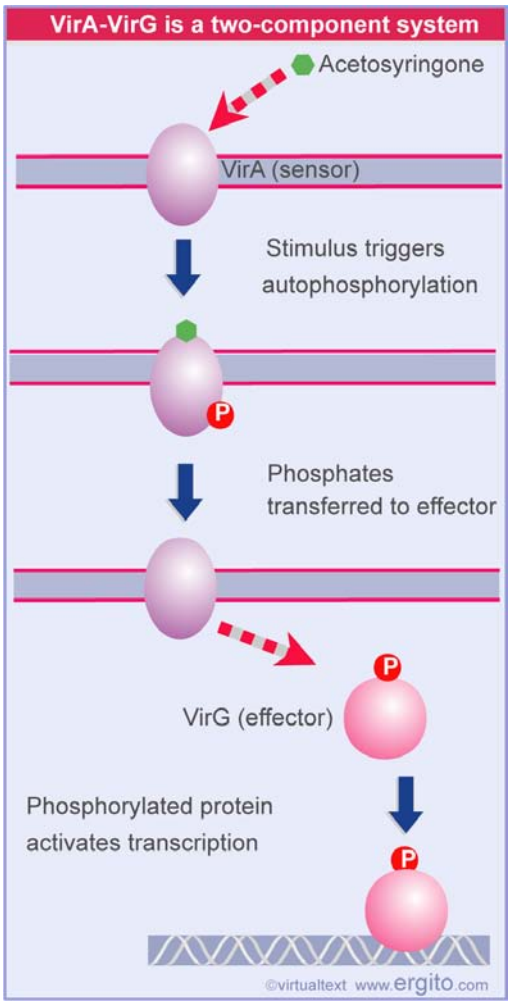

Figure 18.28 The two-component system of VirA-VirG responds to phenolic signals by activating transcription of target genes.
This is a static version of an interactive figure; see http://www.ergito.com/main.jsp?bcs=MBIO.4.18.14 to view properly.

VirA forms a homodimer that is located in the inner membrane; it may respond to the presence of the phenolic compounds in the periplasmic space. Exposure to these compounds causes VirA to become autophosphorylated on histidine. The phosphate group is then transferred to an Asp residue in VirG. The phosphorylated VirG binds to promoters of the *virB,C,D,E* genes to activate transcription. When *virG* is activated, its transcription is induced from a new startpoint, different from that used for constitutive expression, with the result that the amount of VirG protein is increased.

Of the other *vir* loci, *virD* is the best characterized. The *virD* locus has 4 open reading frames. Two of the proteins coded at *virD*, VirD1 and VirD2, provide an endonuclease that initiates the transfer process by nicking T-DNA at a specific site.

REARRANGEMENT OF DNA

4.18.15 Transfer of T-DNA resembles bacterial conjugation

Key Concepts

- T-DNA is generated when a nick at the right boundary creates a primer for synthesis of a new DNA strand.
- The preexisting single-strand that is displaced by the new synthesis is transferred to the plant cell nucleus.
- Transfer is terminated when DNA synthesis reaches a nick at the left boundary.
- The T-DNA is transferred as a complex of single-stranded DNA with the VirE2 single strand-binding protein.
- The single stranded T-DNA is converted into double-stranded DNA and integrated into the plant genome.
- The mechanism of integration is not known. T-DNA can be used to transfer genes into a plant nucleus.

The transfer process actually selects the T-region for entry into the plant. **Figure 18.29** shows that the T-DNA of a nopaline plasmid is demarcated from the flanking regions in the Ti plasmid by repeats of 25 bp, which differ at only two positions between the left and right ends. When T-DNA is integrated into a plant genome, it has a well-defined right junction, which retains 1-2 bp of the right repeat. The left junction is variable; the boundary of T-DNA in the plant genome may be located at the 25 bp repeat or at one of a series of sites extending over ~100 bp within the T-DNA. Sometimes multiple tandem copies of T-DNA are integrated at a single site.

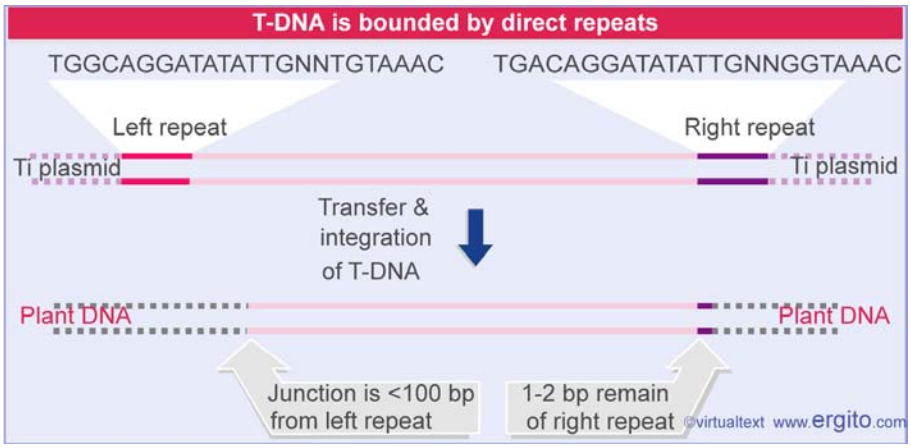

Figure 18.29 T-DNA has almost identical repeats of 25 bp at each end in the Ti plasmid. The right repeat is necessary for transfer and integration to a plant genome. T-DNA that is integrated in a plant genome has a precise junction that retains 1-2 bp of the right repeat, but the left junction varies and may be up to 100 bp short of the left repeat.

A model for transfer is illustrated in **Figure 18.30**. A nick is made at the right 25 bp repeat. It provides a priming end for synthesis of a DNA single strand. Synthesis of the new strand displaces the old strand, which is used in the transfer process. Transfer is terminated when DNA synthesis reaches a nick at the left repeat. This model explains why the right repeat is essential, and it accounts for the polarity of the process. If the left repeat fails to be nicked, transfer could continue farther along the Ti plasmid.

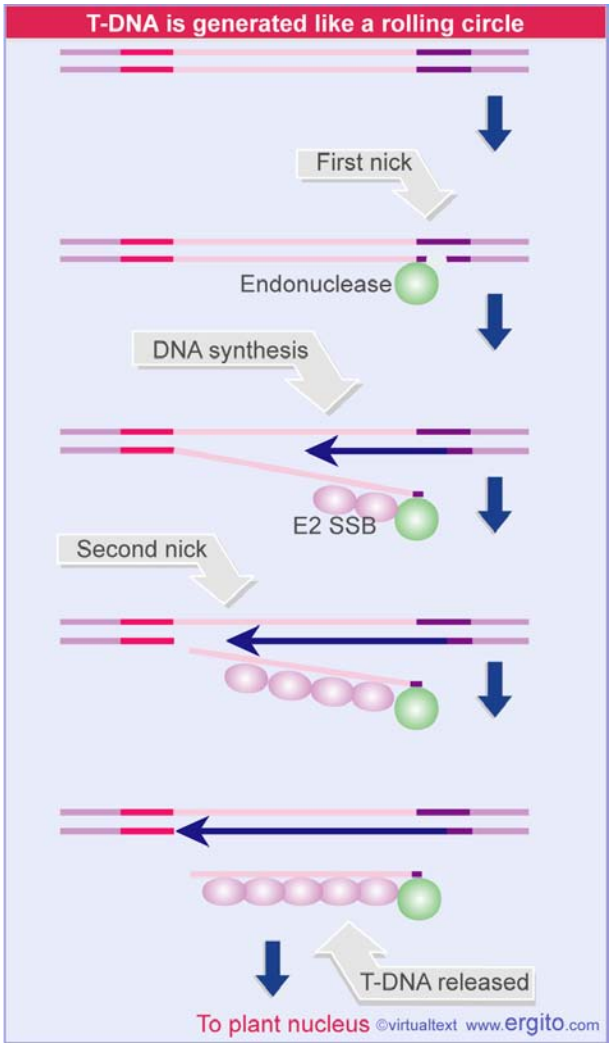

Figure 18.30 T-DNA is generated by displacement when DNA synthesis starts at a nick made at the right repeat. The reaction is terminated by a nick at the left repeat.

The transfer process involves production of a single molecule of single-stranded DNA in the infecting bacterium. It is transferred in the form of a DNA-protein complex, sometimes called the T-complex. The DNA is covered by the VirE2 single-strand binding protein, which has a nuclear localization signal and is responsible for transporting T-DNA into the plant cell nucleus. A single molecule of the D2 subunit of the endonuclease remains bound at the 5′ end. The *virB* operon codes for 11 products that are involved in the transfer reaction (586; for review see 180; 1133).

Outside T-DNA, but immediately adjacent to the right border, is another short sequence, called *overdrive*, which greatly stimulates the transfer process. Overdrive functions like an enhancer: it must lie on the same molecule of DNA, but enhances the efficiency of transfer even when located several thousand base pairs away from the border. VirC1, and possibly VirC2, may act at the overdrive sequence.

Octopine plasmids have a more complex pattern of integrated T-DNA than nopaline plasmids. The pattern of T-strands is also more complex, and several discrete species can be found, corresponding to elements of T-DNA. This suggests that octopine T-DNA has several sequences that provide targets for nicking and/or termination of DNA synthesis.

This model for transfer of T-DNA closely resembles the events involved in bacterial conjugation, when the *E. coli* chromosome is transferred from one cell to another in single-stranded form. The genes of the *virB* operon are homologous to the *tra* genes of certain bacterial plasmids that are involved in conjugation (see *Molecular Biology 4.13.13 Conjugation transfers single-stranded DNA*). A difference is that the transfer of T-DNA is (usually) limited by the boundary of the left repeat, whereas transfer of bacterial DNA is indefinite.

We do not know how the transferred DNA is integrated into the plant genome. At some stage, the newly generated single strand must be converted into duplex DNA. Circles of T-DNA that are found in infected plant cells appear to be generated by recombination between the left and right 25 bp repeats, but we do not know if they are intermediates. The actual event is likely to involve a nonhomologous recombination, because there is no homology between the T-DNA and the sites of integration.

Is T-DNA integrated into the plant genome as an integral unit? How many copies are integrated? What sites in plant DNA are available for integration? Are genes in T-DNA regulated exclusively by functions on the integrated segment? These questions are central to defining the process by which the Ti plasmid transforms a plant cell into a tumor.

What is the structure of the target site? Sequences flanking the integrated T-DNA tend to be rich in A·T base pairs (a feature displayed in target sites for some transposable elements). The sequence rearrangements that occur at the ends of the integrated T-DNA make it difficult to analyze the structure. We do not know whether the integration process generates new sequences in the target DNA comparable to the target repeats created in transposition.

T-DNA is expressed at its site of integration. The region contains several transcription units, each probably containing a gene expressed from an individual promoter. Their functions are concerned with the state of the plant cell, maintaining its tumorigenic properties, controlling shoot and root formation, and suppressing differentiation into other tissues. None of these genes is needed for T-DNA transfer.

The Ti plasmid presents an interesting organization of functions. Outside the T-region, it carries genes needed to initiate oncogenesis; at least some are concerned with the transfer of T-DNA, and we should like to know whether others function in the plant cell to affect its behavior at this stage. Also outside the T-region are the genes that enable the *Agrobacterium* to catabolize the opine that the transformed plant cell will produce. Within the T-region are the genes that control the transformed state of the plant, as well as the genes that cause it to synthesize the opines that will benefit the *Agrobacterium* that originally provided the T-DNA.

As a practical matter, the ability of *Agrobacterium* to transfer T-DNA to the plant

genome makes it possible to introduce new genes into plants. Because the transfer/integration and oncogenic functions are separate, it is possible to engineer new Ti plasmids in which the oncogenic functions have been replaced by other genes whose effect on the plant we wish to test. The existence of a natural system for delivering genes to the plant genome should greatly facilitate genetic engineering of plants.

Reviews

180. Zambryski, P. (1988). *Basic processes underlying Agrobacterium-mediated DNA transfer to plant cells.* Annu. Rev. Genet. 22, 1-30.

1133. Zambryski, P. (1989). *Agrobacterium-plant cell DNA transfer*. Mobile DNA, 309-334.

References

586. Weaver, J. C., Vaughan, T. E., and Astumian, R. D. (2000). *Biological sensing of small field differences by magnetically sensitive chemical reactions.* Nature 405, 707-709.

REARRANGEMENT OF DNA

4.18.16 DNA amplification generates extra gene copies

Key Terms

Amplification refers to the production of additional copies of a chromosomal sequence, found as intrachromosomal or extrachromosomal DNA.

Methotrexate is a drug that inhibits the enzyme DHFR (dihydrofolate reductase).

A **homogeneously staining region (HSR)** is produced by the tandem amplification of a chromosomal sequence. As a result, it does not have a banded pattern.

Double-minute chromosomes are extrachromosomal elements formed by amplification of DHFR genes in response to methotrexate treatment. They are large enough to be visible in the light microscope.

Key Concepts

- Eukaryotic cells acquire resistance to methotrexate by amplifying the number of *dhfr* genes.
- The initial step of amplification is the generation of extrachromosomal DNA molecules containing tandem repeats of the DHFR locus.
- The extrachromosomal DNA forms double minute chromosomes that are lost frequently at division.
- Stable resistant lines are generated by generation of amplified copies in the chromosome.
- It is not known whether stable lines arise by *de novo* amplification *in situ* or by insertion of extrachromosomal amplified sequences.

The eukaryotic genome has the capacity to accommodate additional sequences of either exogenous or endogenous origin. Endogenous sequences may be produced by **amplification** of an existing sequence. The additional sequences often take the form of a tandem array, containing many copies of a repeating unit. A gene that is contained within the repeating unit is not necessarily expressed in every copy, but expression tends to increase with the number of copies.

A tandem array of multiple copies may exist in either of two forms in a cell. If it takes the form of an extrachromosomal unit, it is inherited in an irregular manner: there is no equivalent in animal cells to the ability of a bacterial plasmid to be segregated evenly at cell division, so the entire unit is lost at a high frequency. If the array is integrated into the genome, however, it becomes a component of the genotype, and is inherited like any other genomic sequence.

Amplification of endogenous sequences is provoked by selecting cells for resistance to certain agents. The best-characterized example of amplification results from the

addition of **methotrexate** (mtx) to certain cultured cell lines. This reagent blocks folate metabolism. Resistance to it is conferred by mutations that change the activity of the enzyme dihydrofolate reductase (DHFR). As an alternative to change in the enzyme itself, the amount of enzyme may be increased. The cause of this increase is an amplification of the number of *dhfr* structural genes. Amplification occurs at a frequency greater than the spontaneous point mutation rate, generally ranging from 10^{-4}-10^{-6}. Similar events now have been observed in >20 other genes.

A common feature in most of these systems is that highly resistant cells are not obtained in a single step, but instead appear when the cells are adapted to gradually increasing doses of the toxic reagent. So gene amplification may require several stages. Amplification generally occurs at only one of the two *dhfr* alleles; and increased resistance to methotrexate is accomplished by further increases in the degree of amplification at this locus (for review see 184; 185).

The number of *dhfr* genes in a cell line resistant to methotrexate varies from 40-400, depending on the stringency of the selection and the individual cell line. The *mtx*^{r} lines fall into two classes, distinguished by their response when the selective pressure for high levels of DHFR activity is relieved by growth in the absence of methotrexate. The basis for the difference is illustrated in **Figure 18.31**.

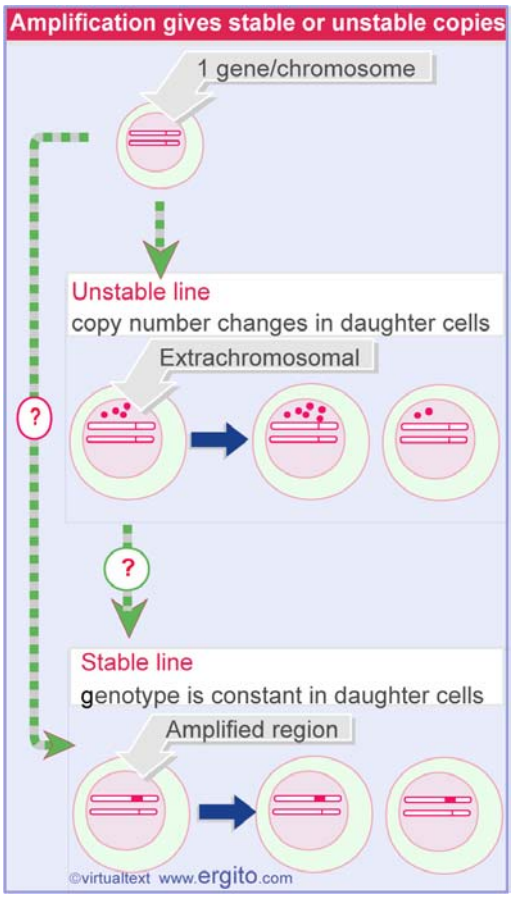

Figure 18.31 The *dhfr* gene can be amplified to give unstable copies that are extrachromosomal (double minutes) or stable (chromosomal). Extrachromosomal copies arise at early times.

- In *stable* lines, the amplified genes are retained, because they reside on the chromosome, at the site usually occupied by the single *dhfr* gene. Usually the other chromosome retains its normal single copy of *dhfr*.
- In *unstable* lines, the amplified genes are at least partially lost when the selective pressure is released, because the amplified genes exist as an extrachromosomal array.

Gene amplification has a visible effect on the chromosomes. In stable lines, the *dhfr* locus can be visualized in the form of a **homogeneously staining region** (**HSR**). An example is shown in **Figure 18.32**. The HSR takes its name from the presence of an additional region that lacks any chromosome bands after treatments such as G-banding. This change suggests that some region of the chromosome between bands has undergone an expansion.

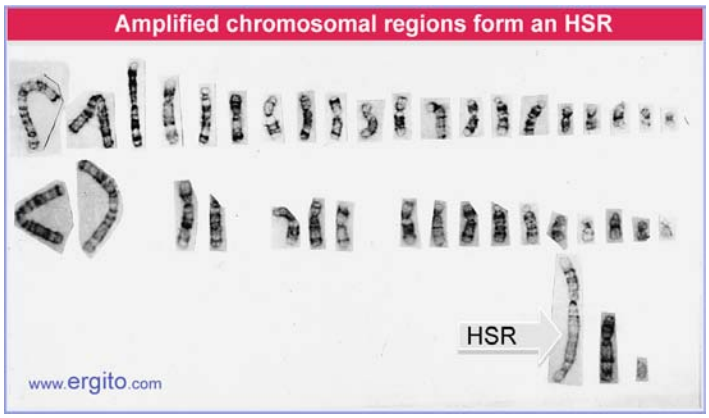

Figure 18.32 Amplified copies of the *dhfr* gene produce a homogeneously staining region (*HSR*) in the chromosome. Photograph kindly provided by Robert Schimke.

In unstable cell lines, no change is seen in the chromosomes carrying *dhfr*. However, large numbers of elements called **double-minute chromosomes** are visible, as can be seen in **Figure 18.33**. In a typical cell line, each double-minute carries 2-4 *dhfr* genes. The double minutes appear to be self-replicating; but they lack centromeres. As a result, they do not attach to the mitotic spindle and therefore segregate erratically, frequently being lost from the daughter cells. Notwithstanding their name, the actual status of the double minutes is regarded as extrachromosomal.

Figure 18.33 Amplified extrachromosomal *dhfr* genes take the form of double-minute chromosomes, as seen in the form of the small white dots. Photograph kindly provided by Robert Schimke.

The irregular inheritance of the double minutes explains the instability of methotrexate resistance in these lines. Double minutes are lost continuously during cell divisions; and in the presence of methotrexate, cells with reduced numbers of *dhfr* genes will die. Only those cells that have retained a sufficient number of double minutes will appear in the surviving population.

The presence of the double minutes reduces the rate at which the cells proliferate. So when the selective pressure is removed, cells lacking the amplified genes have an advantage; they generate progeny more rapidly and soon take over the population. This explains why the amplified state is retained in the cell line only so long as cells

are grown in the presence of methotrexate.

Because of the erratic segregation of the double minutes, increases in the copy number can occur relatively quickly as cells are selected at each division for progeny that have gained more than their fair share of the *dhfr* genes. Cells with greater numbers of copies are found in response to increased levels of methotrexate. The behavior of the double-minutes explains the stepwise evolution of the *mtx*^{r} condition and the incessant fluctuation in the level of *dhfr* genes in unstable lines.

Both stable and unstable lines are found after long periods of selection for methotrexate resistance. What is the initial step in gene amplification? After short periods of selection, most or all of the resistant cells are unstable. The formation of extrachromosomal copies clearly is a more frequent event than amplification within the chromosome. At very early times in the process, amplified *dhfr* genes can be found as (small) extrachromosomal units before double minutes or any change in chromosomes can be detected. This suggests that the acquisition of resistance is most often due to generation of extrachromosomal repeats.

The amplified region is longer than the *dhfr* gene itself. The gene has a length of ~31 kb, but the average length of the repeated unit is 500-1000 kb in the chromosomal HSR. The extent of the amplified region is different in each cell line. The amount of DNA contained in a double minute seems to lie in a range of 100-1000 kb.

How do the extrachromosomal copies arise? We know that their generation occurs without loss of the original chromosomal copy. There are two general possibilities. Additional cycles of replication could be initiated in the vicinity of the *dhfr* gene, followed by nonhomologous recombination between the copies. Or the process could be initiated by nonhomologous recombination between alleles. The extra copies could be released from the chromosome, possibly by some recombination-like event. Depending on the nature of this event, it could generate an extrachromosomal DNA molecule containing one or several copies. If the double minutes contain circular DNA, recombination between them in any case is likely to generate multimeric molecules.

Some information about the events involved in perpetuating the double minutes is given by an unstable cell line whose amplified genes code for a mutant DHFR enzyme. The mutant enzyme is not present in the original (diploid) cell line (so the mutation must have arisen at some point during the amplification process). Despite variations in the number of amplified genes, these cells display only the mutant enzyme. So the wild-type chromosomal genes cannot be continuously generating large numbers of double minutes anew, because these amplified copies would produce normal enzyme.

Once amplified extrachromosomal genes have arisen, therefore, changes in the state of the cell are mediated through these genes and not through the original chromosomal copies. When methotrexate is removed, the cell line loses its double minutes in the usual way. On re-exposure to the reagent, normal genes are amplified to give a new population of double minutes. This shows that none of the extrachromosomal copies of the mutant gene had integrated into the chromosome.

Another striking implication of these results is that the double minutes of the mutant

line carried only mutant genes – so if there is more than one *dhfr* gene per double minute, all must be of the mutant type. This suggests that multicopy double minutes can be generated from individual extrachromosomal genes.

A major question has been whether amplified chromosomal copies arise by integration of the extrachromosomal copies or by an independent mechanism. We do not know whether intrachromosomal amplification simply proceeds less often as a *de novo* step or requires extrachromosomal amplification to occur as an intermediate step. The form taken by the amplified genes is influenced by the cell genotype; some cell lines tend to generate double minutes, while others more readily display the HSR configuration.

The type of amplification event also depends upon the particular locus that is involved. Another case of amplification is provided by resistance to an inhibitor of the enzyme transcarbamylase, which occurs by amplification of the CAD gene. (CAD is a protein which has the first three enzymatic activities of the pathway for UMP synthesis.) Amplified CAD DNA is always found within the chromosome. In this case, the amplified genes are found in the form in the form of several dispersed amplified regions, often involving more than one chromosome.

Reviews

184. Schimke, R. T (1981). *Chromosomal and extrachromosomal localization of amplified DHFR genes in cultured mammalian cells*. Cold Spring Harbor Symp. Quant. Biol. 45, 785-797.

185. Stark, G. R. and Wahl, G. M. (1984). *Gene amplification*. Annu. Rev. Biochem. 53, 447-491.

REARRANGEMENT OF DNA

4.18.17 Transfection introduces exogenous DNA into cells

Key Terms

Transfection of eukaryotic cells is the acquisition of new genetic markers by incorporation of added DNA.

Transient transfectants have foreign DNA in an unstable, i.e. extrachromosomal, form.

An **integrant (Stable transfectant)** is a cell line in which a gene introduced by transfection has become integrated into the genome.

Key Concepts

- DNA that is transfected into a eukaryotic cells forms a large repeating unit of many head to tail tandem repeats.
- The transfected unit is unstable unless it becomes integrated into a host chromosome.
- Genes carried by the transfected DNA can be expressed.

The procedure for introducing exogenous donor DNA into recipient cells is called **transfection**. Transfection experiments began with the addition of preparations of metaphase chromosomes to cell suspensions. The chromosomes are taken up rather inefficiently by the cells and give rise to unstable variants at a low frequency. Intact chromosomes rarely survive the procedure; the recipient cell usually gains a fragment of a donor chromosome (which is unstable because it lacks a centromere). Rare cases of stable lines may have resulted from integration of donor material into a resident chromosome.

Similar results are obtained when purified DNA is added to a recipient cell preparation. However, with purified DNA it is possible to add particular sequences instead of relying on random fragmentation of chromosomes. Transfection with DNA yields stable as well as unstable lines, with the former relatively predominant. (These experiments are directly analogous to those performed in bacterial transformation, but are described as transfection because of the historical use of "transformation" to describe changes that allow unrestrained growth of eukaryotic cells.)

Unstable transfectants (sometimes called **transient transfectants**) reflect the survival of the transfected DNA in extrachromosomal form; stable lines result from integration into the genome. The transfected DNA can be expressed in both cases. However, the low frequencies of transfection make it necessary to use donor markers whose presence in the recipient cells can be selected for. Most transfection experiments have used markers representing readily assayed enzymatic functions, but, in principle, any marker that can be selected can be assayed. This allows the

isolation of genes responsible for morphological phenomena. Most notably, transfected cells can be selected for acquisition of the transformed (tumorigenic) phenotype. This type of protocol has led to the isolation of several cellular *onc* genes (see *Molecular Biology 6.30.9 Ras oncogenes can be detected in a transfection assay*).

Cotransfection with more than one marker has proved informative about the events involved in transfection and extends the range of questions that we can ask with this technique. A common marker used in such experiments is the *tk* gene, coding for the enzyme thymidine kinase, which catalyzes an essential step in the provision of thymidine triphosphate as a precursor for DNA synthesis.

When tk^- cells are transfected with a DNA preparation containing both a purified tk^+gene and the ϕX174 genome, all the tk^+ transformants have both donor sequences. This is a useful observation, because it allows unselected markers to be introduced routinely by cotransfection with a selected marker.

The arrangement of *tk* and ϕX174 sequences is different in each transfected line, but remains the same during propagation of that line. Often multiple copies of the donor sequences are present, the number varying with the individual line. Revertants lose the ϕX174 sequences together with *tk* sequences. Amplification of transfected sequences under selective pressure results in the increase of copy number of all donor sequences *pari passu*. So the two types of donor sequence become physically linked during transfection and suffer the same fate thereafter.

To perform a transfection experiment, the mass of DNA added to the recipient cells is increased by including an excess of "carrier DNA," a preparation of some other DNA (often from salmon sperm). Transfected cells prove to have sequences of the carrier DNA flanking the selected sequences on either side. Transfection therefore appears to be mediated by a large unit, consisting of a linked array of all sequences present in the donor preparation.

Since revertants for the selected marker lose all of this material, it seems likely that the transfected cell gains only a single large unit. The unit is formed by a concatemeric linkage of donor sequences in a reaction that is rapid relative to the other events involved in transfection. This transfecting package is ~1000 kb in length.

Because of the size of the donor unit, we cannot tell from blotting experiments whether it is physically linked to recipient chromosomal DNA (the relevant end fragments are present in too small a relative proportion). It seems plausible that the first stage is the establishment of an unstable extrachromosomal unit, followed by the acquisition of stability via integration.

In situ hybridization can be used to show that transfected cells have donor material integrated into the resident chromosomes. Any given cell line has only a single site of integration; but the site is different in each line. Probably the selection of a site for integration is a random event; sometimes it is associated with a gross chromosomal rearrangement.

The sites at which exogenous material becomes integrated usually do not appear to

have any sequence relationship to the transfected DNA. The integration event involves a nonhomologous recombination between the mass of added DNA and a random site in the genome. The recombination event may be provoked by the introduction of a double-strand break into the chromosomal DNA, possibly by the action of DNA repair enzymes that are induced by the free ends of the exogenous DNA. **Integrants** produced by the integration event are stable, and are therefore more useful than transient transfectants for experiments that rely on the expression of the transfected gene (for review see 183).

Reviews

183. Pellicer, A. (1980). *Altering genotype and phenotype by DNA-mediated gene transfer*. Science 209, 1414-1422.

REARRANGEMENT OF DNA

4.18.18 Genes can be injected into animal eggs

Key Terms

Transgenic animals are created by introducing DNA prepared in test tubes into the germline. The DNA may be inserted into the genome or exist in an extrachromsomal structure.

Key Concepts

- DNA that is injected into animal eggs can integrate into the genome.
- Usually a large array of tandem repeats integrates at a single site.
- Expression of the DNA is variable and may be affected by the site of integration and other epigenetic effects.

An exciting development of transfection techniques is their application to introduce genes into animals. An animal that gains new genetic information from the addition of foreign DNA is described as **transgenic**. The approach of directly injecting DNA can be used with mouse eggs, as shown in **Figure 18.34**. Plasmids carrying the gene of interest are injected into the germinal vesicle (nucleus) of the oocyte or into the pronucleus of the fertilized egg. The egg is implanted into a pseudopregnant mouse. After birth, the recipient mouse can be examined to see whether it has gained the foreign DNA, and, if so, whether it is expressed. (For a description of the technique see *Great Experiments 11.3 Transgenic mice: Expression of foreign genes in animals*.)

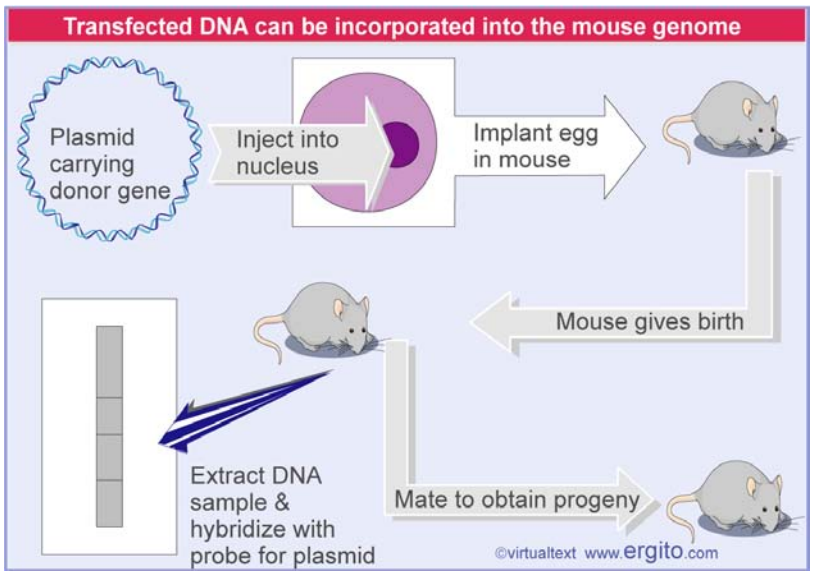

Figure 18.34 Transfection can introduce DNA directly into the germ line of animals

The first questions we ask about any transgenic animal are how many copies it has of the foreign material, where these copies are located, and whether they are present in the germline and inherited in a Mendelian manner. The usual result of such experiments is that a reasonable minority (say ~15%) of the injected mice carry the transfected sequence. Usually, multiple copies of the plasmid appear to have been integrated in a tandem array into a single chromosomal site. The number of copies varies from 1-150. They are inherited by the progeny of the injected mouse as expected of a Mendelian locus.

An important issue that can be addressed by experiments with transgenic animals concerns the independence of genes and the effects of the region within which they reside. If we take a gene, including the flanking sequences that contain its known regulatory elements, can it be expressed independently of its location in the genome? In other words, do the regulatory elements function independently, or is gene expression in addition controlled by other effects, for example, location in an appropriate chromosomal domain?

Are transfected genes expressed with the proper developmental specificity? The general rule now appears to be that there is a reasonable facsimile of proper control: the transfected genes are generally expressed in appropriate cells and at the usual time. There are exceptions, however, in which a transfected gene is expressed in an inappropriate tissue.

In the progeny of the injected mice, expression of the donor gene is extremely variable; it may be extinguished entirely, reduced somewhat, or even increased. Even in the original parents, the level of gene expression does not correlate with the number of tandemly integrated genes. Probably only some of the genes are active. In addition to the question of how many of the gene copies are capable of being activated, a parameter influencing regulation could be the relationship between the gene number and the regulatory proteins: a large number of promoters could dilute out any regulator proteins present in limiting amounts.

What is responsible for the variation in gene expression? One possibility that has often been discussed for transfected genes (and which applies also to integrated retroviral genomes) is that the site of integration is important. Perhaps a gene is expressed if it integrates within an active domain, but not if it integrates in another area of chromatin. Another possibility is the occurrence of epigenetic modification; for example, changes in the pattern of methylation might be responsible for changes in activity. Alternatively, the genes that happened to be active in the parents may have been deleted or amplified in the progeny.

A particularly striking example of the effects of an injected gene is provided by a strain of transgenic mice derived from eggs injected with a fusion consisting of the MT promoter linked to the rat growth hormone structural gene. Growth hormone levels in some of the transgenic mice were several hundred times greater than normal. The mice grew to nearly twice the size of normal mice, as can be seen from **Figure 18.35**.

Figure 18.35 A transgenic mouse with an active rat growth hormone gene (left) is twice the size of a normal mouse (right). Photograph kindly provided by Ralph Brinster.

The introduction of oncogene sequences can lead to tumor formation. Transgenic mice containing the SV40 early coding region and regulatory elements express the viral genes for large T and small t antigens only in some tissues, most often brain, thymus, and kidney. (The T/t antigens are alternatively spliced proteins coded by the early region of the virus; they have the ability to transform cultured cells to a tumorigenic phenotype; see *Molecular Biology 6.30.5 Early genes of DNA transforming viruses have multifunctional oncogenes*.) The transgenic mice usually die before reaching 6 months, as the result of developing tumors in the brain; sometimes tumors are found also in thymus and kidney. Different oncogenes may be used to generate mice developing various cancers, thus making possible a range of model systems. For example, introduction of the *myc* gene under control of an active promoter causes the appearance of adenocarcinomas and other tumors.

Can defective genes be replaced by functional genes in the germline using transgenic techniques? One successful case is represented by a cure of the defect in the hypogonadal mouse. The *hpg* mouse has a deletion that removes the distal part of the gene coding for the polyprotein precursor to GnRH (gonadotropin-releasing hormone) and GnRH-associated peptide (GAP). As a result, the mouse is infertile.

When an intact *hpg* gene is introduced into the mouse by transgenic techniques, it is expressed in the appropriate tissues. **Figure 18.36** summarizes experiments to introduce a transgene into *hpg*/*hpg* homozygous mutant mice. The resulting mice are normal. This provides a striking demonstration that expression of a transgene under normal regulatory control can be indistinguishable from the behavior of the normal allele.

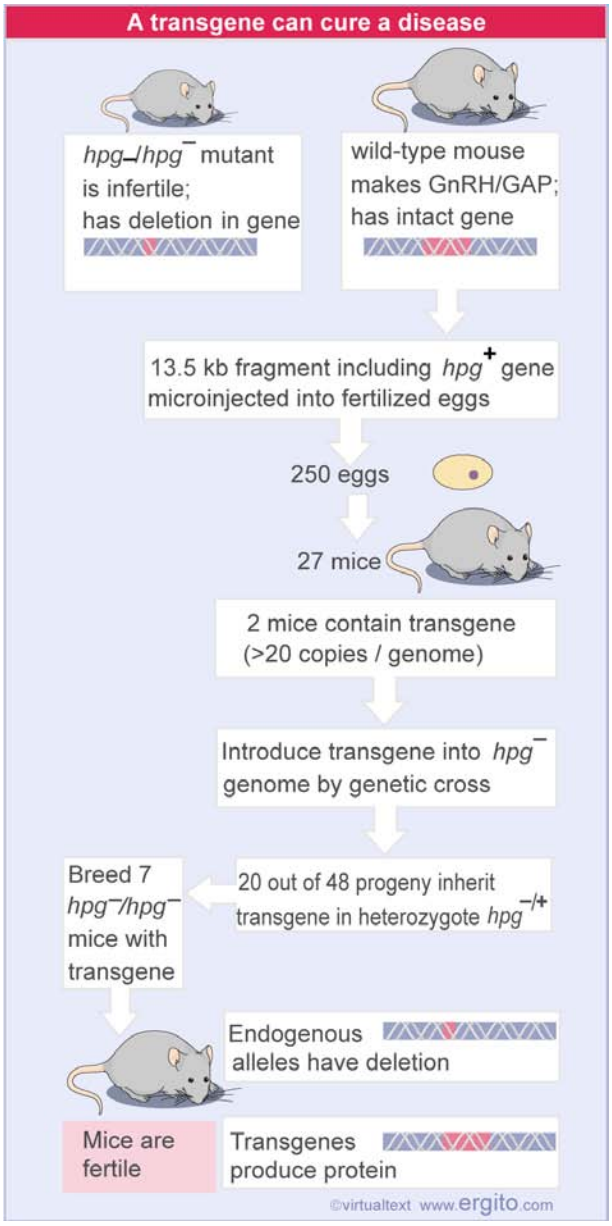

Figure 18.36 Hypogonadism of the *hpg* mouse can be cured by introducing a transgene that has the wild-type sequence.

Impediments to using such techniques to cure genetic defects at present are that the transgene must be introduced into the germline of the preceding generation, the ability to express a transgene is not predictable, and an adequate level of expression of a transgene may be obtained in only a small minority of the transgenic animals. Also, the large number of transgenes that may be introduced into the germline, and their erratic expression, could pose problems for the animal in cases in which overexpression of the transgene was harmful.

In the *hpg* murine experiments, for example, only 2 out of 250 eggs mice injected with intact hpg genes gave rise to transgenic mice. Each transgenic animal contained >20 copies of the transgene. Only 20 of the 48 offspring of the transgenic mice

retained the transgenic trait. When inherited by their offspring, however, the transgene(s) could substitute for the lack of endogenous *hpg* genes. Gene replacement via a transgene is therefore effective only under restricted conditions.

The disadvantage of direct injection of DNA is the introduction of multiple copies, their variable expression, and often difficulty in cloning the insertion site because sequence rearrangements may have been generated in the host DNA. An alternative procedure is to use a retroviral vector carrying the donor gene. A single proviral copy inserts at a chromosomal site, without inducing any rearrangement of the host DNA. It is possible also to treat cells at different stages of development, and thus to target a particular somatic tissue; however, it is difficult to infect germ cells.

REARRANGEMENT OF DNA

4.18.19 ES cells can be incorporated into embryonic mice

Key Concepts

- ES (embryonic carcinoma) cells that are injected into a mouse blastocyst generate descendant cells that become part of a chimeric adult mouse.
- When the ES cells contribute to the germline, the next generation of mice may be derived from the ES cell.
- Genes can be added to the mouse germline by transfecting them into ES cells before the cells are added to the blastocyst.

A powerful technique for making transgenic mice takes advantage of embryonic stem (ES) cells, which are derived from the mouse blastocyst (an early stage of development, which precedes implantation of the egg in the uterus). **Figure 18.37** illustrates the principles of this technique.

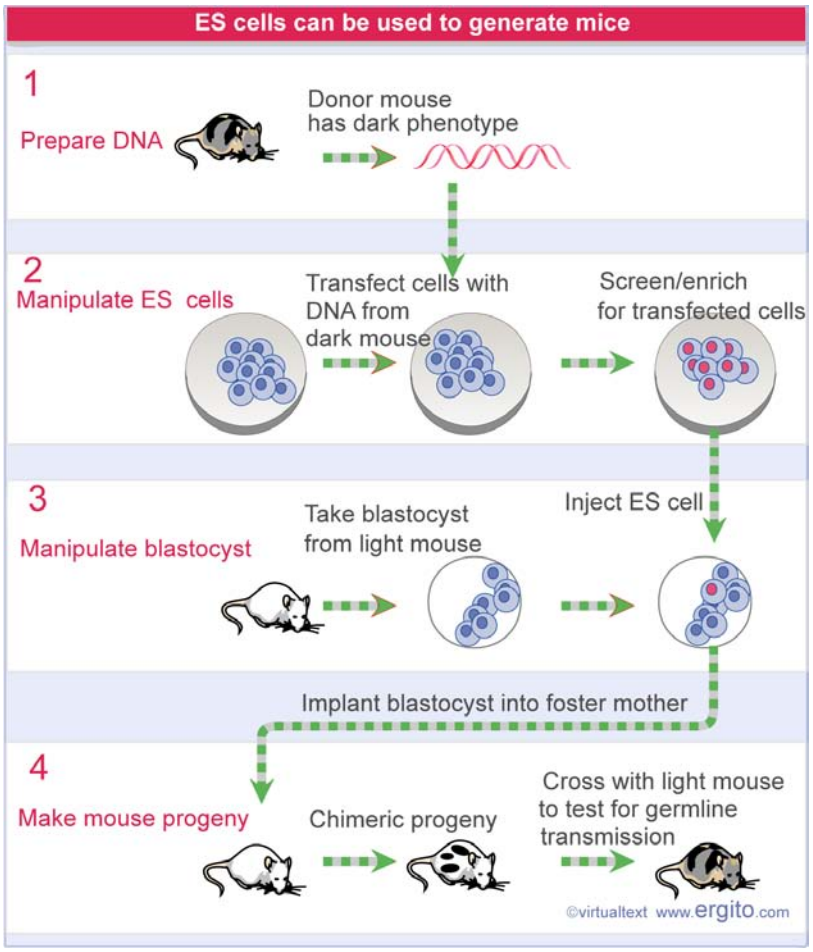

Figure 18.37 ES cells can be used to generate mouse chimeras, which breed true for the transfected DNA when the ES cell contributes to the germ line.

ES cells are transfected with DNA in the usual way (most often by microinjection or electroporation). By using a donor that carries an additional sequence such as a drug resistance marker or some particular enzyme, it is possible to select ES cells that have obtained an integrated transgene carrying any particular donor trait. An alternative is to use PCR technology to assay the transfected ES cells for successful integration of the donor DNA. By such means, a population of ES cells is obtained in which there is a high proportion carrying the marker.

These ES cells are then injected into a recipient blastocyst. The ability of the ES cells to participate in normal development of the blastocyst forms the basis of the technique. The blastocyst is implanted into a foster mother, and in due course develops into a *chimeric* mouse. Some of the tissues of the chimeric mice will be derived from the cells of the recipient blastocyst; other tissues will be derived from the injected ES cells. The proportion of tissues in the adult mouse that are derived from cells in the recipient blastocyst and from injected ES cells varies widely in individual progeny; if a visible marker (such as coat color gene) is used, areas of tissue representing each type of cell can be seen.

To determine whether the ES cells contributed to the germline, the chimeric mouse is

crossed with a mouse that lacks the donor trait. Any progeny that have the trait must be derived from germ cells that have descended from the injected ES cells. By this means, an entire mouse has been generated from an original ES cell (for review see 186)!

Reviews

186. Jaenisch, R. (1988). *Transgenic animals*. Science 240, 1468-1474.

REARRANGEMENT OF DNA

4.18.20 Gene targeting allows genes to be replaced or knocked out

Key Concepts

- An endogenous gene can be replaced by a transfected gene using homologous recombination.
- The occurrence of a homologous recombination can be detected by using two selectable markers, one of which is incorporated with the integrated gene, the other of which is lost when recombination occurs.

A further development of these techniques makes it possible to obtain homologous recombinants (see *Great Experiments 3.4 Gene targeting: Altering the genome in mice*). A particular use of homologous recombination is to disrupt endogenous genes, as illustrated in **Figure 18.38**. A wild-type gene is modified by interrupting an exon with a marker sequence; most often the *neo* gene that confers resistance to the drug G418 is used. Also, another marker is added on one side of the gene; for example, the TK gene of the herpes virus. When this DNA is introduced into an ES cell, it may be inserted into the genome by either nonhomologous or homologous recombination. A nonhomologous recombination inserts the whole unit, including the flanking TK sequence. But a homologous recombination requires two exchanges, as a result of which the flanking TK sequence is lost. Cells in which a homologous recombination has occurred can therefore be selected by the gain of *neo* resistance and absence of TK activity (which can be selected because TK causes sensitivity to the drug gancyclovir). If it is not convenient to use a selectable marker such as TK, cells can simply be screened by PCR assays for the absence of flanking DNA. The frequency of homologous recombination is ~10^{-7}, and probably represents <1% of all recombination events (for review see 187).

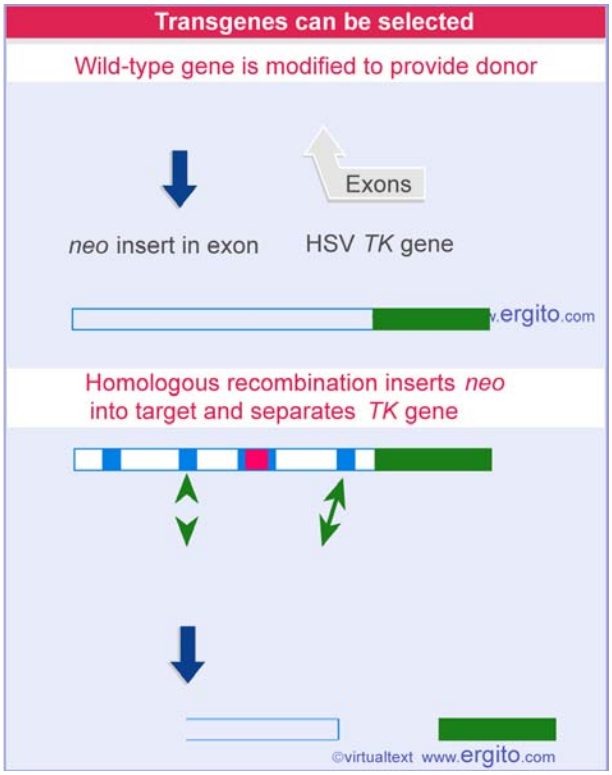

Figure 18.38 A transgene containing *neo* within an exon and TK downstream can be selected by resistance to G418 and loss of TK activity.

The presence of the *neo* gene in an exon disrupts transcription, and thereby creates a null allele. A particular target gene can therefore be "knocked out" by this means; and once a mouse with one null allele has been obtained, it can be bred to generate the homozygote. This is a powerful technique for investigating whether a particular gene is essential, and what functions in the animal are perturbed by its loss.

A sophisticated method for introducing new DNA sequences has been developed with *D. melanogaster* by taking advantage of the P element. The protocol is illustrated in **Figure 18.39**. A defective P element carrying the gene of interest is injected together with an intact P element into preblastoderm embryos. The intact P element provides a transposase that recognizes not only its own ends but also those of the defective element. As a result, either or both elements may be inserted into the genome (587).

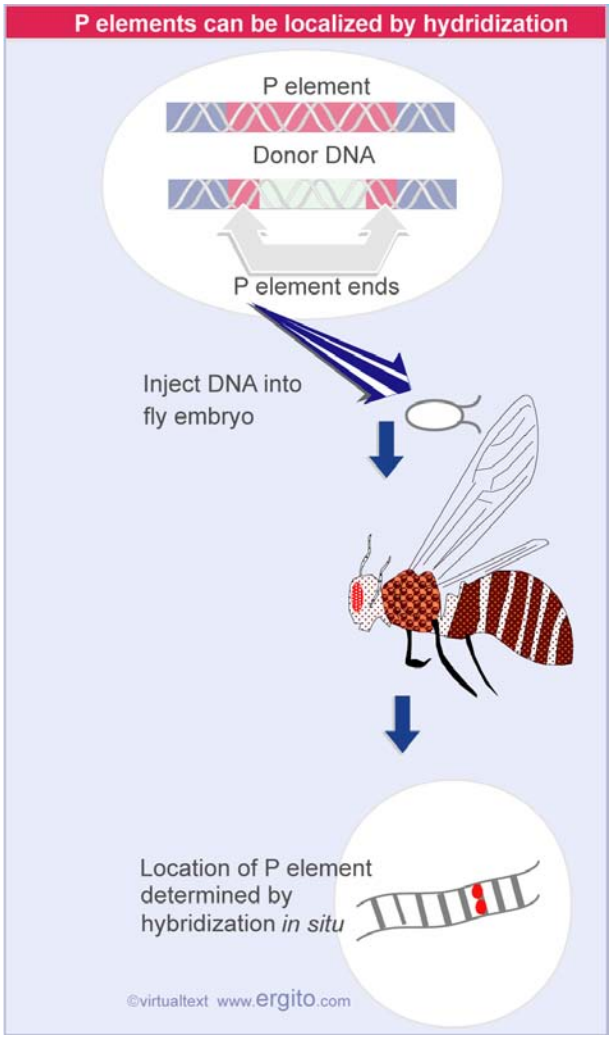

Figure 18.39 Transgenic flies that have a single, normally expressed copy of a gene can be obtained by injecting *D. melanogaster* embryos with an active P element plus foreign DNA flanked by P element ends.

Only the sequences between the ends of the P DNA are inserted; the sequences on either side are not part of the transposable element. An advantage of this technique is that only a single element is inserted in any one event, so the transgenic flies usually carry only one copy of the foreign gene, a great aid in analyzing its behavior.

Several genes that have been introduced in this way all show the same behavior. They are expressed only in the appropriate tissues and at the proper times during development, irrespective of the site of integration. So in *D. melanogaster*, all the information needed to regulate gene expression may be contained within the gene locus itself, and can be relatively impervious to external influence.

With these experiments, we see the possibility of extending from cultured cells to animals the option of examining the regulatory features. The ability to introduce DNA into the genotype allows us to make changes in it, to add new genes that have had particular modifications introduced *in vitro*, or to inactivate existing genes. So it

becomes possible to delineate the features responsible for tissue-specific gene expression. Ultimately we may expect routinely to replace defective genes in the genotype in a targeted manner.

Reviews

187. Capecchi, M. R. (1989). *Altering the genome by homologous recombination*. Science 244, 1288-1292.

References

587. Spradling, A. C. and Rubin, G. M. (1982). *Transposition of cloned P elements into Drosophila germline chromosomes*. Science 218, 341-353.

REARRANGEMENT OF DNA

4.18.21 Summary

Yeast mating type is determined by whether the *MAT* locus carries the *a* or α sequence. Expression in haploid cells of the sequence at *MAT* leads to expression of genes specific for the mating type and to repression of genes specific for the other mating type. Both activation and repression are achieved by control of transcription, and require factors that are not specific for mating type as well as the products of *MAT*. The functions that are activated in either mating type include secretion of the appropriate pheromone and expression on the cell surface of the receptor for the opposite type of pheromone. Interaction between pheromone and receptor on cells of either mating type activates a G protein on the membrane, and sets in train a common pathway that prepares cells for sporulation. Diploid cells do not express mating-type functions.

Additional, silent copies of the mating-type sequences are carried at the loci *HML* α and *HMRa*. They are repressed by the actions of the *sir* loci. Cells that carry the HO endonuclease display a unidirectional transfer process in which the sequence at *HML* α replaces an *a* sequence at *MAT*, or the sequence at *HMRa* replaces an α sequence at *MAT*. The endonuclease makes a double-strand break at *MAT*, and a free end invades either *HML* α or *HMRa*. *MAT* initiates the transfer process, but is the recipient of the new sequence. The HO endonuclease is transcribed in mother cells but not daughter cells, and is under cell-cycle control. So switching is detected only in the products of a division, and the mating type has been switched in both daughter cells.

Trypanosomes carry >1000 sequences coding for varieties of the surface antigen. Only a single VSG is expressed in one cell, from an active site located near a telomere. The active site is localized in a special extranucleolar body, where the VSG gene is transcribed by RNA polymerase I. The VSG may be changed by substituting a new coding sequence at the active site via a gene conversion process, or by switching the site of expression to another telomere. Switches in expression occur every 10^4-10^6 divisions.

Agrobacteria induce tumor formation in wounded plant cells. The wounded cells secrete phenolic compounds that activate *vir* genes carried by the Ti plasmid of the bacterium. The *vir* gene products cause a single strand of DNA from the T-DNA region of the plasmid to be transferred to the plant cell nucleus. Transfer is initiated at one boundary of T-DNA, but ends at variable sites. The single strand is converted into a double strand and integrated into the plant genome. Genes within the T-DNA transform the plant cell, and cause it to produce particular opines (derivatives of arginine). Genes in the Ti plasmid allow *Agrobacterium* to metabolize the opines produced by the transformed plant cell. T-DNA has been used to develop vectors for transferring genes into plant cells.

Endogenous sequences may become amplified in cultured cells. Exposure to methotrexate leads to the accumulation of cells that have additional copies of the *dhfr* gene. The copies may be carried as extrachromosomal arrays in the form of double-minute "chromosomes," or they may be integrated into the genome at the site

of one of the *dhfr* alleles. Double-minute chromosomes are unstable, and disappear from the cell line rapidly in the absence of selective pressure. The amplified copies may originate by additional cycles of replication that are associated with recombination events.

New sequences of DNA may be introduced into a cultured cell by transfection or into an animal egg by microinjection. The foreign sequences may become integrated into the genome, often as large tandem arrays. The array appears to be inherited as a unit in a cultured cell. The sites of integration appear to be random. A transgenic animal arises when the integration event occurs into a genome that enters the germ-cell lineage. A transgene or transgenic array is inherited in Mendelian manner, but the copy number and activity of the gene(s) may change in the progeny. Often a transgene responds to tissue- and temporal regulation in a manner that resembles the endogenous gene. Using conditions that promote homologous recombination, an inactive sequence can be used to replace a functional gene, thus creating a null locus. Transgenic mice can be obtained by injecting recipient blastocysts with ES cells that carry transfected DNA.