Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002.

General Recombination

In the two preceding sections, we discussed the mechanisms that allow the DNA sequences in cells to be maintained from generation to generation with very little change. However, it is also clear that these DNA sequences can occasionally be rearranged. The particular combination of genes present in any individual genome, as well as the timing and the level of expression of these genes, is often altered by such DNA rearrangements. In a population, this type of genetic variation is crucial to allow organisms to evolve in response to a changing environment. The DNA rearrangements are caused by a set of mechanisms that are collectively called **genetic recombination**. Two broad classes are commonly recognized—*general recombination* and *site-specific recombination*. In this part of the chapter we discuss the first of these two mechanisms; in the next part, we consider the second mechanism.

In **general recombination** (also known as *homologous recombination),* genetic exchange takes place between a pair of homologous DNA sequences. These are usually located on two copies of the same chromosome, although other types of DNA molecules that share the same nucleotide sequence can also participate. The general recombination reaction is essential for every proliferating cell, because accidents occur during nearly every round of DNA replication that interrupt the replication fork and require general recombination mechanisms to repair. The details of the intimate interplay between replication and recombination are still incompletely understood, but they include using variations of the homologous end-joining reaction (see Figure 5-53) to restart replication forks that have run into a break in the parental DNA template.

General recombination is also essential for the accurate chromosome segregation that occurs during meiosis in fungi, plants, and animals (see Figure 20-11). The crossing-over of chromosomes that results causes bits of genetic information to be exchanged to create new combinations of DNA sequences in each chromosome. The evolutionary benefit of this type of gene mixing is apparently so great that the reassortment of genes by general recombination is not confined to multicellular organisms; it is also widespread in single-celled organisms.

The central features that lie at the heart of the general recombination mechanism seem to be the same in all organisms. Most of what we know about the biochemistry of genetic recombination was originally derived from studies of bacteria, especially of *E. coli* and its viruses, as well as from experiments with simple eucaryotes such as yeasts. For these organisms with short generation times and relatively small genomes, it was possible to isolate a large set of mutants with defects in their recombination processes. The identification of the protein altered in each mutant then allowed the collection of proteins that catalyze general recombination to be identified and characterized. More recently, close relatives of these proteins have been discovered and extensively characterized in *Drosophila,* mice, and humans as well.

General Recombination Is Guided by Base-pairing Interactions Between Two Homologous DNA Molecules

The abundant general recombination observed in meiosis has the following characteristics: (1) Two homologous DNA molecules that were originally part of different chromosomes “cross over;” that is, their double helices break and the two broken ends join to their opposite partners to re-form two intact double helices, each composed of parts of the two initial DNA molecules (Figure 5-54). (2) The site of exchange (that is, where a *red* double helix is joined to a *green* double helix in Figure 5-54) can occur anywhere in the homologous nucleotide sequences of the two participating DNA molecules. (3) At the site of exchange, a strand of one DNA molecule has become base-paired to a strand of the second DNA molecule to create a *heteroduplex joint* that

links the two double helices (Figure 5-55). This heteroduplex region can be thousands of base pairs long; we explain later how it forms. (4) No nucleotide sequences are altered at the site of exchange; some DNA replication usually takes place, but the cleavage and rejoining events occur so precisely that not a single nucleotide is lost or gained. Despite its precision, general recombination creates DNA molecules of novel sequence: the heteroduplex joint can tolerate a small number of mismatched base pairs, and, more importantly, the two DNA molecules that cross over are usually not exactly the same on either side of the joint. As a result, new recombinant DNA molecules (recombinant chromosomes) are generated.

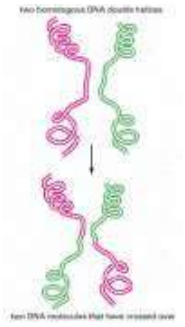


Figure 5-54

General recombination. The breaking and rejoining of two homologous DNA double helices creates two DNA molecules that have "crossed over." In meiosis, this process causes each chromosome in a germ cell to contain a mixture of maternally (more...)

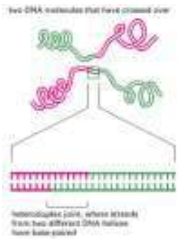


Figure 5-55

A heteroduplex joint. This structure unites two DNA molecules where they have crossed over. Such a joint is often thousands of nucleotides long.

The mechanism of general recombination ensures that two DNA double helices undergo an exchange reaction only if they contain an extensive region of sequence similarity (homology). The formation of a long heteroduplex joint requires such homology because it involves a long region of complementary base-pairing between a strand from one of the two original double helices and a complementary strand from the other double helix. But how does this heteroduplex joint arise, and how do the two homologous regions of DNA at the site of crossing-over recognize each other? As we shall see, recognition takes place during a process called **DNA synapsis**, in which base pairs form between complementary strands from the two DNA molecules. This base-pairing is then extended to guide the general recombination process, allowing it to occur only between DNA molecules that contain long regions of matching (or nearly matching) DNA sequence.

Meiotic Recombination Is Initiated by Double-strand DNA Breaks

Extensive base-pair interactions cannot occur between two intact DNA double helices. Thus, the DNA synapsis that is critical for general recombination in meiosis can begin only after a DNA strand from one DNA helix has been exposed and its nucleotides have been made available for pairing with another DNA helix. In the absence of direct experimental evidence, theoretical models were proposed based on the idea that a break needed to be made in just one of the two strands of a DNA helix to produce the exposed DNA strand required for DNA synapsis. This break in the phosphodiester backbone was thought to allow one of the nicked strand ends to separate from its base-paired partner strand, freeing it to form a short heteroduplex with a second intact DNA helix—thereby beginning synapsis. Models of this type are reasonable in theory, and they have been described in textbooks for nearly 30 years.

In the early 1990s, sensitive biochemical techniques became available for determining the actual structure of the recombination intermediates that form in yeast chromosomes at various stages of meiosis. These studies revealed that general recombination is initiated by a special endonuclease that simultaneously cuts *both* strands of the double helix, creating a complete break in the DNA

molecule. The 5′ ends at the break are then chewed back by an exonuclease, creating protruding single-stranded 3′ ends. It is these single strands that search for a homologous DNA helix with which to pair—leading to the formation of a joint molecule between a maternal and a paternal chromosome (Figure 5-56).

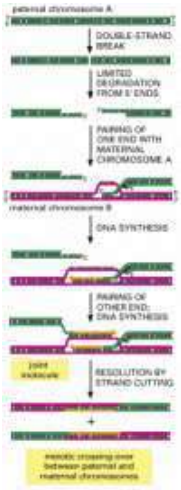


Figure 5-56

General recombination in meiosis. As indicated, the process begins when an endonuclease makes a double-strand break in a chromosome. An exonuclease then creates two protruding 3′ single-stranded ends, which find the homologous region of a second (more...)

In the next section, we begin to explain how a DNA single strand can "find" a homologous double-stranded DNA molecule to begin DNA synapsis.

DNA Hybridization Reactions Provide a Simple Model for the Base-pairing Step in General Recombination

In its simplest form, the type of base-pairing interaction central to the synapsis step of general recombination can be mimicked in a test tube by allowing a DNA double helix to re-form from its separated single strands. This process, called *DNA renaturation* or hybridization, occurs when a rare random collision juxtaposes complementary nucleotide sequences on two matching DNA single strands, allowing the formation of a short stretch of double helix between them. This relatively slow helix nucleation step is followed by a very rapid "zippering" step, as the region of double helix is extended to maximize the number of base-pairing interactions (Figure 5-57).

Figure 5-57

DNA hybridization. DNA double helices re-form from their separated strands in a reaction that depends on the random collision of two complementary DNA strands. The vast majority of such collisions are not productive, as shown on the left, but a few result (more...)

Formation of a new double helix in this way requires that the annealing strands be in an open, unfolded conformation. For this reason, *in vitro* hybridization reactions are performed at either high temperature or in the presence of an organic solvent such as formamide; these conditions "melt out" the short hairpin helices that result from the base-pairing interactions that occur within a single strand that folds back on itself. Most cells cannot survive such harsh conditions and instead use a single-strand DNA-binding (SSB) protein (see p. 246) to melt out the hairpin helices and help anneal their complementary single strands. This protein is essential for DNA replication (as described earlier) as well as for general recombination; it binds tightly and cooperatively to the sugar-phosphate backbone of all single-stranded DNA regions of DNA, holding them in an extended conformation with the bases exposed (see Figures 5-17 and 5-18). In this extended conformation, a DNA single strand can base-pair efficiently either with a nucleoside triphosphate molecule (in DNA replication) or with a complementary section of another DNA single strand (as part of a genetic recombination process).

The partner that a DNA single-strand needs to find in the synapsis step of general recombination is a DNA double helix, rather than a second single strand of DNA (see Figure 5-56). In the next section we see how the critical event that allows DNA hybridization to begin during recombination—the initial invasion of a single-stranded DNA into a DNA double helix—is achieved by the cell.

The RecA Protein and its Homologs Enable a DNA Single Strand to Pair with a Homologous Region of DNA Double Helix

General recombination is more complex than the simple hybridization reactions just described involving single-stranded DNA, and it requires several types of specialized proteins. In particular, the *E. coli* RecA protein has a central role in the recombination between chromosomes; it and its homologs in yeast, mice, and humans make synapsis possible (Figure 5-58).

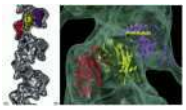


Figure 5-58

The structure of the RecA and Rad51 protein–DNA filaments. (A) The Rad51 protein bound to a DNA single strand. Rad51 is a human homolog of the bacterial RecA protein; three successive monomers in this helical filament are colored. (B) A short (more...)

Like a single-strand DNA-binding protein, the RecA type of protein binds tightly and in long cooperative clusters to single-stranded DNA to form a nucleoprotein filament. Because each RecA monomer has more than one DNA-binding site, a RecA filament can hold a single strand and a double helix together. This allows it to catalyze a multistep DNA synapsis reaction between a DNA double helix and a homologous region of single-stranded DNA. The region of homology is identified before the duplex DNA target has been opened up, through a three-stranded intermediate in which the DNA single strand forms transient base pairs with bases that flip out from the helix in the major groove of the double-stranded DNA molecule (Figure 5-59). This reaction begins the pairing shown previously in Figure 5-56, and it thereby initiates the exchange of strands between two recombining DNA double helices.

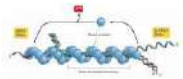
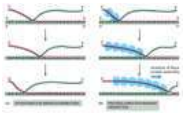


Figure 5-59

DNA synapsis catalyzed by the RecA protein. *In vitro* experiments show that several types of complex are formed between a DNA single strand covered with RecA protein *(red)* and a DNA double helix *(green)*. First a non-base-paired complex is formed, which (more...)

Once DNA synapsis has occurred, the short heteroduplex region where the strands from two different DNA molecules have begun to pair is enlarged through a process called *branch migration*. Branch migration can take place at any point where two single DNA strands with the same sequence are attempting to pair with the same complementary strand; in this reaction, an unpaired region of one of the single strands displaces a paired region of the other single strand, moving the branch point without changing the total number of DNA base pairs. Although spontaneous branch migration can occur, it proceeds equally in both directions, so it makes little progress and is unlikely to complete recombination efficiently (Figure 5-60A). The RecA protein catalyzes unidirectional branch migration, readily producing a region of heteroduplex DNA that is thousands of base pairs long (Figure 5-60B).

Figure 5-60



Two types of DNA branch migration observed in experiments Spontaneous branch migration is a back-and-forth, random-wa therefore makes little progress over long distances. (B) RecA- branch migration proceeds (more...)

The catalysis of directional branch migration depends on a further property of the RecA protein. In addition to having two DNA-binding sites, the RecA protein is a DNA-dependent ATPase, with an additional site for binding and hydrolyzing ATP. The protein associates much more tightly with DNA when it has ATP bound than when it has ADP bound. Moreover, new RecA molecules with ATP bound are preferentially added at one end of the RecA protein filament, and the ATP is then hydrolyzed to ADP. The RecA protein filaments that form on DNA may therefore share some of the dynamic properties displayed by the cytoskeletal filaments formed from actin or tubulin (discussed in Chapter 16); an ability of the protein to "treadmill" unidirectionally along a DNA strand, for example, could drive the branch migration reaction shown in Figure 5-60B.

There Are Multiple Homologs of the RecA Protein in Eucaryotes, Each Specialized for a Specific Function

When one compares the proteins that catalyze the basic genetic functions in eucaryotes with those in bacteria such as *E. coli*, one generally finds that evolutionarily related proteins are present that catalyze similar reactions. In many cases, however, multiple eucaryotic homologs take the place of a particular bacterial protein, each specialized for a specific aspect of the bacterial protein's function.

This generalization applies to the *E. coli* RecA protein: humans and mice contain at least seven RecA homologs. Each homolog is presumed to have special catalytic activities and its own set of accessory proteins. The *Rad51 protein* is a particularly important RecA homolog in yeast, mice, and humans; it catalyzes a synaptic reaction between a DNA single strand and a DNA double helix in experiments *in vitro*. Genetic studies in which the Rad51 protein is mutated suggest that this protein is critical for the health of all three organisms, being required to repair replication forks that have been accidentally broken during the normal course of each S phase. Its proper function requires multiple accessory proteins. Two of these, the *Brca1* and *Brca2* proteins, were first discovered because mutations in their genes are inherited in a subset of human families with a greatly increased frequency of breast cancer. Whereas the removal of the Rad51 protein kills a cell, less drastic changes in its function caused by an alteration in such an accessory protein is thought to lead to an accumulation of DNA damage that often, in a small proportion of cells, gives rise to a cancer (see Figure 23-11).

Different RecA homologs in eucaryotes are specialized for meiosis, or for other unique types of DNA synaptic events that are less well understood. It is likely that each eucaryotic RecA homolog loads onto a DNA single strand to begin a general recombination event only when a particular DNA structure or cell condition allows the protein to bind there.

General Recombination Often Involves a Holliday Junction

The synapsis that exchanges the first single strand between two different DNA double helices is presumed to be the slow and difficult step in a general recombination event (see Figure 5-56). After this step, extending the region of pairing and establishing further strand exchanges between the two DNA helices is thought to proceed rapidly. In most cases, a key recombination intermediate, the *Holliday junction* (also called a *cross-strand exchange)* forms as a result.

In a Holliday junction, the two homologous DNA helices that have initially paired are held together by the reciprocal exchange of two of the four strands present, one originating from each of the helices. As shown in Figure 5-61A, a Holliday junction can be considered to contain two

pairs of strands: one pair of crossing strands and one pair of noncrossing strands. The structure can *isomerize,* however, by undergoing a series of rotational movements, catalyzed by specialized proteins, to form a more open structure in which both pairs of strands occupy equivalent positions (Figures 5-61B and 5-62). This structure can, in turn, isomerize to a conformation that closely resembles the original junction, except that the crossing strands have been converted into noncrossing strands, and vice versa (Figure 5-61C).

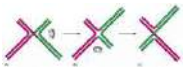


Figure 5-61

A Holliday junction and its isomerization. As described in the text, the synapsis step in general recombination is catalyzed by a RecA type of protein bound to a DNA single strand. This step is often followed by a reciprocal exchange of strands between (more...)

Figure 5-62

Electron micrograph of a Holliday junction. This view of the junction corresponds to the open structure illustrated in Figure 5-61B. (Courtesy of Huntington Potter and David Dressler.)

Once the Holliday junction has formed an open structure, a special set of proteins can engage with the junction: one of these proteins uses the energy of ATP hydrolysis to move the crossover point (the point at which the two DNA helices are joined) rapidly along the two helices, extending the region of heteroduplex DNA (Figure 5-63).

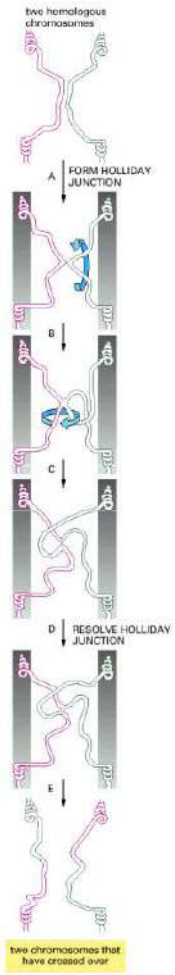
Figure 5-63

Enzyme-catalyzed double branch migration at a Holliday junction. In *E. coli*, a tetramer of the RuvA protein *(green)* and two hexamers of the RuvB protein *(pale gray)* bind to the open form of the junction. The RuvB protein uses the energy of ATP hydrolysis (more...)

To regenerate two separate DNA helices, and thus end the exchange process, the strands connecting the two helices in a Holliday junction must eventually be cut, a process referred to as *resolution*. There are two ways in which a Holliday junction can be resolved. In one, the original pair of crossing strands is cut (the invading, or inside, strands in Figure 5-61A). In this case, the two original DNA helices separate from each other nearly unaltered, exchanging the single-stranded DNA that formed the heteroduplex. In the other way, the original pair of noncrossing strands is cut (the inside strands in Figure 5-61C). Now the outcome is far more profound: two recombinant chromosomes are formed, having reciprocally exchanged major segments of double-stranded DNA with each other through a crossover event (Figure 5-64).

Figure 5-64

The resolution of a Holliday junction to produce crossed-over chromosomes. In this example, homologous regions of a *red* and a *green* chromosome have formed a Holliday junction by exchanging two strands. Cutting these two strands would terminate the exchange (more...)



Genetic analyses reveal that heteroduplex regions of several thousand base pairs are readily formed during recombination. As described next, the processing of these heteroduplexes—which generally consist of nearly identical paired complementary strands—can further change the information in each resulting DNA helix.

General Recombination Can Cause Gene Conversion

In sexually reproducing organisms, it is a fundamental law of genetics that each parent makes an equal genetic contribution to an offspring, which inherits one complete set of genes from the father and one complete set from the mother. Thus, when a diploid cell undergoes meiosis to produce four haploid cells (discussed in Chapter 20), exactly half of the genes in these cells should be maternal (genes that the diploid cell inherited from its father) and the other half paternal (genes that the diploid cell inherited from its father). In some organisms (fungi, for example), it is possible to recover and analyze all four of the haploid gametes produced from a single cell by meiosis. Studies in such organisms have revealed rare cases in which the standard rules of genetics have been violated. Occasionally, for example, meiosis yields three copies of the maternal version of the gene and only one copy of the paternal allele (alternative versions of the same gene are called alleles). This phenomenon is known as gene conversion (Figure 5-65). Gene conversion often occurs in association with homologous genetic recombination events in meiosis (and more rarely in mitosis), and it is believed to be a straightforward consequence of the mechanisms of general recombination and DNA repair. Genetic studies show that only small sections of DNA typically undergo gene conversion, and in many cases only a part of a gene is changed.

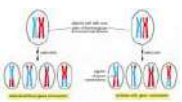


Figure 5-65

Gene conversion in meiosis. As described in Chapter 20, meiosis is the process through which a diploid cell gives rise

to four haploid cells. Germ cells (eggs and sperm, for example) are produced by meiosis.

In the process of gene conversion, DNA sequence information is transferred from one DNA helix that remains unchanged (a donor sequence) to another DNA helix whose sequence is altered (an acceptor sequence). There are several different ways this might happen, all of which involve the following two processes: (1) a homologous recombination event that juxtaposes two homologous DNA double helices, and (2) a limited amount of localized DNA synthesis, which is necessary to create an extra copy of one allele. In the simplest case, a general recombination process forms a heteroduplex joint (see Figure 5-55), in which the two paired DNA strands are not identical in sequence and therefore contain some mismatched base pairs. If the mispaired nucleotides in one of the two strands are recognized and removed by the DNA repair enzyme that catalyzes mismatch repair, an extra copy of the DNA sequence on the opposite strand is produced (Figure 5-66). The same gene conversion process can occur without crossover events, since it simply requires that a single DNA strand invade a double helix to form a short heteroduplex region. The latter type of gene conversion is thought to be responsible for the unusually facile transfer of genetic information that is often observed between the different gene copies in a tandem array of repeated genes.

Figure 5-66

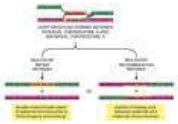
Gene conversion by mismatch correction. In this process, heteroduplex joints are formed at the sites of the crossing-over between homologous maternal and paternal chromosomes. If the maternal and paternal DNA sequences are slightly different, the heteroduplex (more...)

General Recombination Events Have Different Preferred Outcomes in Mitotic and Meiotic Cells

We have seen that meiotic recombination starts with a very bold stroke—the breakage of both strands of the double helix in one of the recombining chromosomes. How does the meiotic process that follows differ from the mechanism, also based on general recombination, that cells use for the precise repair of the accidental double-strand breaks that occur in chromosomes (the homologous end-joining reaction in Figure 5-53)? In both cases, the two new chromosome ends produced by a double-strand break are subjected to a degradative process, which exposes a single strand with an overhanging 3′ end. Moreover, in both cases, this strand seeks out a region of unbroken DNA double helix with the same nucleotide sequence and undergoes a synaptic reaction with it that is catalyzed by a RecA type of protein.

For double-strand break repair, DNA synthesis extends the invading 3′ end by thousands of nucleotides, using one of the strands of the recipient DNA helix as a template. If the second broken end becomes similarly engaged in the synaptic reaction, a joint molecule will be formed (see Figure 5-56). Depending on subsequent events, the final outcome can either be restoration of the two original DNA helices with repair of the double-strand break (the predominant reaction in mitotic cells), or a crossover event that leaves heteroduplex joints holding two different DNA helices together (the predominant reaction in meiotic cells). It is thought that the crossover events are created by a set of specific proteins that guide these reactions cells undergoing meiosis. These proteins not only ensure that a joint molecule with two Holliday junctions is formed but also cause a different pair of strands at each of the two junctions, thereby causing a crossover event (Figure 5-67).

Figure 5-67



The different resolutions of a general recombination intermedi meiotic cells. As shown previously in Figure 5-56, general rec begins when a double-strand break is generated in one double followed by DNA degradation (more...)

With either outcome of general recombination, the DNA synthesis involved converts some of the genetic information at the site of the double-stranded break to that of the homologous chromosome. If these regions represent different alleles of the same gene, the nucleotide sequence in the broken helix is converted to that of the unbroken helix, causing a gene conversion. The yeast *Saccharomyces cerevisiae* exploits the gene conversion that accompanies double-strand break repair to switch from one mating type to another (discussed in Chapter 7). In this case, a double-strand break is intentionally induced by cleavage of a specific DNA sequence at the yeast *mating type locus* by an enzyme called HO endonuclease. After DNA degradation at the site of the break has removed the old sequence, the missing genetic information is restored by a synapsis of the broken ends with a "mating-type cassette" DNA sequence of the opposite mating type (a or α), followed by local DNA synthesis in the manner previously indicated to reseal the broken region of the chromosome. In fact, it is through a detailed study of this precisely positioned form of double-strand break repair that the general mechanism of homologous end-joining was revealed.

Mismatch Proofreading Prevents Promiscuous Recombination Between Two Poorly Matched DNA Sequences

As previously discussed, a critical step in recombination occurs when two DNA strands of complementary sequence pair to form a heteroduplex joint between two double helices. Experiments *in vitro* with purified RecA protein show that pairing can occur efficiently even when the sequences of the two DNA strands do not match well—when, for example, only four out of every five nucleotides on average can form base pairs. If recombination proceeded from these mismatched sequences, it would create havoc in cells, especially in those that contain a series of closely related DNA sequences in their genomes. How do cells prevent crossing over between these sequences?

Although the complete answer is not known, studies with bacteria and yeasts demonstrate that components of the same mismatch proofreading system that removes replication errors (see Figure 5-23) have the additional role of interrupting genetic recombination events between poorly matched DNA sequences. It has long been known, for example, that homologous genes in two closely related bacteria, *E. coli* and *Salmonella typhimurium*, generally will not recombine, even though their nucleotide sequences are 80% identical. However, when the mismatch proofreading system is inactivated by mutation, there is a 1000-fold increase in the frequency of such interspecies recombination events. It is thought that the mismatch proofreading system normally recognizes the mispaired bases in an initial strand exchange, and—if there are a significant number of mismatches—the subsequent steps required to break and rejoin the two paired DNA helices are prevented. This mechanism protects the bacterial genome from the sequence changes that would otherwise be caused by recombination with the foreign DNA molecules that occasionally enter the cell. In vertebrate cells, which contain many closely related DNA sequences, the same type of recombinational proofreading is thought to help prevent promiscuous recombination events that would otherwise scramble the genome (Figure 5-68).

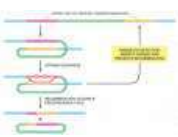


Figure 5-68

The mechanism that prevents general recombination from destabilizing a genome that contains repeated sequences. Studies with bacterial and yeast cells suggest that components

of the mismatch proofreading system, diagrammed previously in Figure 5-23, have (more...)

Summary

General recombination (also called homologous recombination) allows large sections of the DNA double helix to move from one chromosome to another, and it is responsible for the crossing-over of chromosomes that occurs during meiosis in fungi, animals, and plants. General recombination is essential for the maintenance of chromosomes in all cells, and it usually begins with a double-strand break that is processed to expose a single-stranded DNA end. Synapsis between this single strand and a homologous region of DNA double helix is catalyzed by the bacterial RecA protein and its eucaryotic homologs, and it often leads to the formation of a four-stranded structure known as a Holliday junction. Depending on the pattern of strand cuts made to resolve this junction into two separate double helices, the products can be either a precisely repaired double-strand break or two chromosomes that have crossed over.

Because general recombination relies on extensive base-pairing interactions between the strands of the two DNA double helices that recombine, it occurs only between homologous DNA molecules. Gene conversion, the nonreciprocal transfer of genetic information from one chromosome to another, results from the mechanisms of general recombination, which involve a limited amount of associated DNA synthesis.

Bookshelf ID: NBK26898