RECOMBINATION AND REPAIR

4.15.1 Introduction

Key Terms

Homologous recombination (generalized recombination) involves a reciprocal exchange of sequences of DNA, e.g. between two chromosomes that carry the same genetic loci.

Site-specific recombination (Specialized recombination) occurs between two specific sequences, as in phage integration/excision or resolution of cointegrate structures during transposition.

Transposition refers to the movement of a transposon to a new site in the genome.

Copy choice is a type of recombination used by RNA viruses, in which the RNA polymerase switches from one template to another during synthesis.

Evolution could not happen without genetic recombination. If it were not possible to exchange material between (homologous) chromosomes, the content of each individual chromosome would be irretrievably fixed in its particular alleles. When mutations occurred, it would not be possible to separate favorable and unfavorable changes. The length of the target for mutation damage would effectively be increased from the gene to the chromosome. Ultimately a chromosome would accumulate so many deleterious mutations that it would fail to function.

By shuffling the genes, recombination allows favorable and unfavorable mutations to be separated and tested as individual units in new assortments. It provides a means of escape and spreading for favorable alleles, and a means to eliminate an unfavorable allele without bringing down all the other genes with which this allele is associated. This is the basis for natural selection.

Recombination occurs between precisely corresponding sequences, so that not a single base pair is added to or lost from the recombinant chromosomes. Three types of recombination share the feature that the process involves physical exchange of material between duplex DNAs:

- Recombination involving reaction between homologous sequences of DNA is called *generalized* or **homologous recombination**. In eukaryotes, it occurs at meiosis, usually both in males (during spermatogenesis) and females (during oogenesis). We recall that it happens at the "four strand" stage of meiosis, and involves only two of the four strands (see *Molecular Biology 1.1.20 Recombination occurs by physical exchange of DNA*).

- Another type of event sponsors recombination between specific pairs of sequences. This was first characterized in prokaryotes where **specialized recombination**, also known as **site-specific recombination**, is responsible for the integration of phage genomes into the bacterial chromosome. The recombination event involves specific sequences of the phage DNA and bacterial

DNA, which include a short stretch of homology. The enzymes involved in this event act only on the particular pair of target sequences in an intermolecular reaction. Some related intramolecular reactions are responsible during bacterial division for regenerating two monomeric circular chromosomes when a dimer has been generated by generalized recombination. Also in this latter class are recombination events that invert specific regions of the bacterial chromosome.

- A different type of event allows one DNA sequence to be inserted into another without relying on sequence homology. **Transposition** provides a means by which certain elements move from one chromosomal location to another. The mechanisms involved in transposition depend upon breakage and reunion of DNA strands, and thus are related to the processes of recombination (see *Molecular Biology 4.16 Transposons* and *Molecular Biology 4.17 Retroviruses and retroposons*).

- Another type of recombination is used by RNA viruses, in which the polymerase switches from one template to another while it is synthesizing RNA. As a result, the newly synthesized molecule joins sequence information from two different parents. This type of mechanism for recombination is called **copy choice**, and is discussed briefly in *Molecular Biology 4.17.4 Viral DNA is generated by reverse transcription.*

Let's consider the nature and consequences of the generalized and specialized recombination reactions.

Figure 15.1 makes the point that generalized recombination occurs between two homologous DNA duplexes, and can occur at any point along their length. The two chromosomes are cut at equivalent points, and then each is joined to the other to generate reciprocal recombinants. The crossover (marked by the X) is the point at which each becomes joined to the other. There is no change in the overall organization of DNA; the products have the same structure as the parents, and both parents and products are homologous.

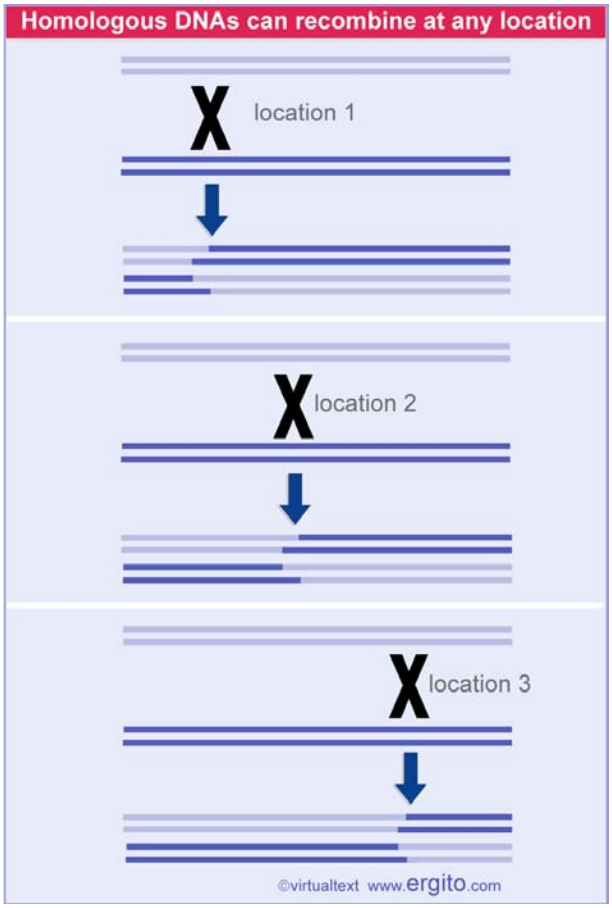

Figure 15.1 Generalized recombination can occur at any point along the lengths of two homologous DNAs.

Specialized recombination occurs only between specific sites. The results depend on the locations of the two recombining sites. **Figure 15.2** shows that an intermolecular recombination between a circular DNA and a linear DNA inserts the circular DNA into the linear DNA. **Figure 15.3** shows that an intramolecular recombination between two sites on a circular DNA releases two smaller circular DNAs. Specialized recombination is often used to make changes such as these in the organization of DNA. The change in organization is a consequence of the locations of the recombining sites. We have a large amount of information about the enzymes that undertake specialized recombination, which are related to the topoisomerases that act to change the supercoiling of DNA in space.

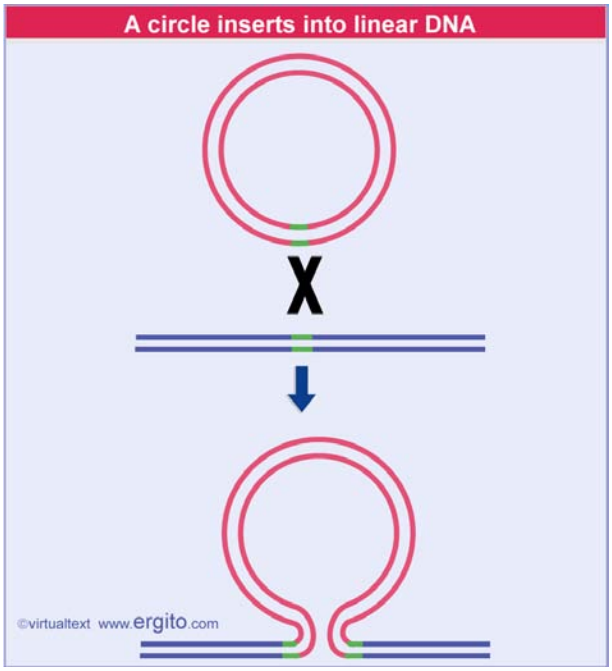

Figure 15.2 Site-specific recombination occurs between two specific sequences (identified in green). The other sequences in the two recombining DNAs are not homologous.

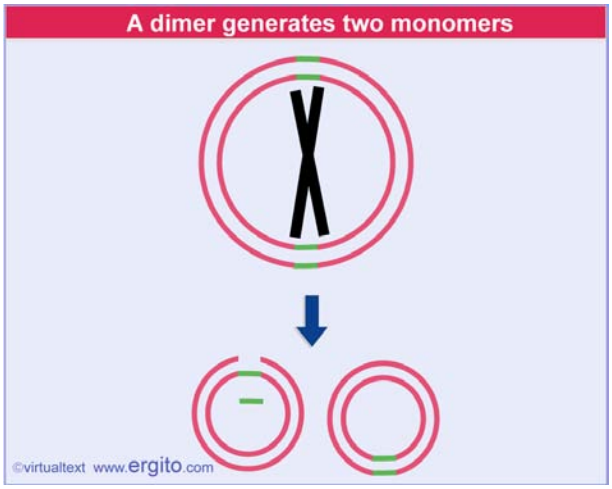

Figure 15.3 Site-specific recombination can be used to generate two monomeric circles from a dimeric circle.

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.4.15.1

RECOMBINATION AND REPAIR

4.15.2 Homologous recombination occurs between synapsed chromosomes

Key Terms

A **bivalent** is the structure containing all four chromatids (two representing each homologue) at the start of meiosis.

Synapsis describes the association of the two pairs of sister chromatids (representing homologous chromosomes) that occurs at the start of meiosis; the resulting structure is called a bivalent.

Chromosome pairing is the coupling of the homologous chromosomes at the start of meiosis.

The **synaptonemal complex** describes the morphological structure of synapsed chromosomes.

Breakage and reunion describes the mode of genetic recombination, in which two DNA duplex molecules are broken at corresponding points and then rejoined crosswise (involving formation of a length of heteroduplex DNA around the site of joining).

A **chiasma** (*pl.* chiasmata) is a site at which two homologous chromosomes appear to have exchanged material during meiosis.

Key Concepts

- Chromosomes must synapse (pair) in order for chiasmata to form where crossing-over occurs.
- We can correlate the stages of meiosis with the molecular events that happen to DNA.

Homologous recombination is a reaction between two duplexes of DNA. Its critical feature is that the enzymes responsible can use any pair of homologous sequences as substrates (although some types of sequences may be favored over others). The frequency of recombination is not constant throughout the genome, but is influenced by both global and local effects. The overall frequency may be different in oocytes and in sperm; recombination occurs twice as frequently in female as in male humans. And within the genome its frequency depends upon chromosome structure; for example, crossing-over is suppressed in the vicinity of the condensed and inactive regions of heterochromatin.

Recombination occurs during the protracted prophase of meiosis. **Figure 15.4** compares the visible progress of chromosomes through the five stages of meiotic prophase with the molecular interactions that are involved in exchanging material between duplexes of DNA.

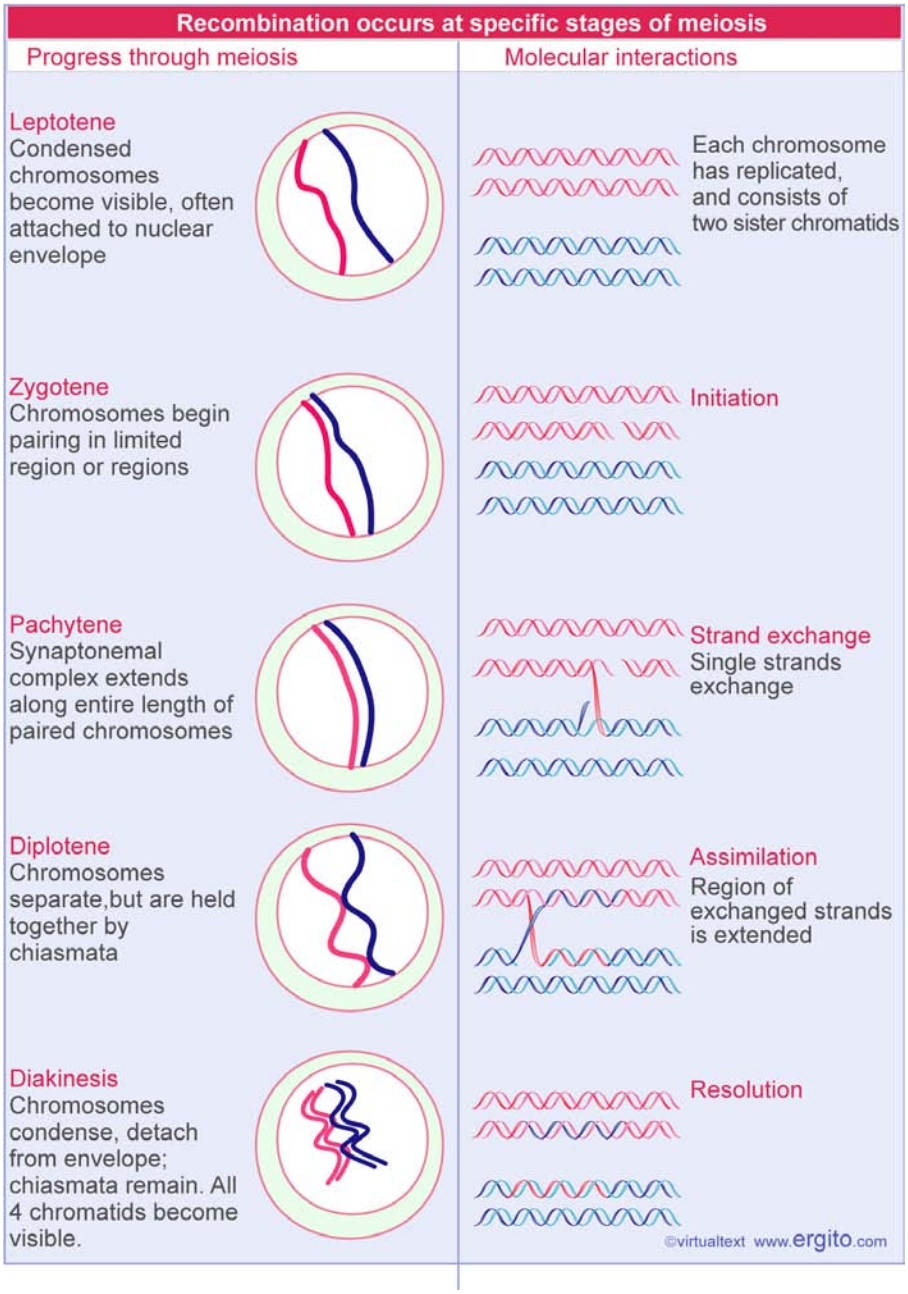

Figure 15.4 Recombination occurs during the first meiotic prophase. The stages of prophase are defined by the appearance of the chromosomes, each of which consists of two replicas (sister chromatids), although the duplicated state becomes visible only at the end. The molecular interactions of any individual crossing-over event involve two of the four duplex DNAs.

The beginning of meiosis is marked by the point at which individual chromosomes become visible. Each of these chromosomes has replicated previously, and consists of two sister chromatids, each of which contains a duplex DNA. The homologous chromosomes approach one another and begin to pair in one or more regions, forming **bivalents**. Pairing extends until the entire length of each chromosome is apposed with its homolog. The process is called **synapsis** or **chromosome pairing**.

When the process is completed, the chromosomes are laterally associated in the form of a **synaptonemal complex**, which has a characteristic structure in each species, although there is wide variation in the details between species.

Recombination between chromosomes involves a physical exchange of parts, usually represented as a **breakage and reunion**, in which two nonsister chromatids (each containing a duplex of DNA) have been broken and then linked each with the other. When the chromosomes begin to separate, they can be seen to be held together at discrete sites, the **chiasmata**. The number and distribution of chiasmata parallel the features of genetic crossing-over. Traditional analysis holds that a chiasma represents the crossing-over event (see **Figure 1.32**). The chiasmata remain visible when the chromosomes condense and all four chromatids become evident.

What is the molecular basis for these events? Each sister chromatid contains a single DNA duplex, so each bivalent contains 4 duplex molecules of DNA. Recombination requires a mechanism that allows the duplex DNA of one sister chromatid to interact with the duplex DNA of a sister chromatid from the other chromosome. It must be possible for this reaction to occur between any pair of corresponding sequences in the two molecules in a highly specific manner that allows material to be exchanged with precision at the level of the individual base pair.

We know of only one mechanism for nucleic acids to recognize one another on the basis of sequence: complementarity between single strands. The figure shows a general model for the involvement of single strands in recombination. The first step in providing single strands is to make a break in each DNA duplex. Then one or both of the strands of that duplex can be released. If (at least) one strand displaces the corresponding strand in the other duplex, the two duplex molecules will be specifically connected at corresponding sequences. If the strand exchange is extended, there can be more extensive connection between the duplex. And by exchanging both strands and later cutting them, it is possible to connect the parental duplex molecules by means of a crossover that corresponds to the demands of a breakage and reunion.

We cannot at this juncture relate these molecular events rigorously with the changes that are observed at the level of the chromosomes. There is no detailed information about the molecular events involved in recombination in higher eukaryotic cells (in which meiosis has been most closely observed). However, recently the isolation of mutants in yeast has made it possible to correlate some of the molecular steps with approximate stages of meiosis. Detailed information about the recombination process is available in bacteria, in which molecular activities are known that cause genetic exchange between duplex molecules. However, the bacterial reaction involves interaction between restricted regions of the genome, rather than an entire pairing of genomes. The synapsis of eukaryotic chromosomes remains the most difficult stage to explain at the molecular level.

RECOMBINATION AND REPAIR

4.15.3 Breakage and reunion involves heteroduplex DNA

Key Terms

A **joint molecule** is a pair of DNA duplexes that are connected together through a reciprocal exchange of genetic material.

A **recombinant joint** is the point at which two recombining molecules of duplex DNA are connected (the edge of the heteroduplex region).

Heteroduplex DNA (Hybrid DNA) is generated by base pairing between complementary single strands derived from the different parental duplex molecules; it occurs during genetic recombination.

Branch migration describes the ability of a DNA strand partially paired with its complement in a duplex to extend its pairing by displacing the resident strand with which it is homologous.

Resolution occurs by a homologous recombination reaction between the two copies of the transposon in a cointegrate. The reaction generates the donor and target replicons, each with a copy of the transposon.

A **Holliday** structure is an intermediate structure in homologous recombination, where the two duplexes of DNA are connected by the genetic material exchanged between two of the four strands, one from each duplex. A joint molecule is said to be resolved when nicks in the structure restore two separate DNA duplexes.

Splice recombinant DNA results from a Holliday junction being resolved by cutting the non-exchanged strands. Both strands of DNA before the exchange point come from one chromosome; the DNA after the exchange point come from the homologous chromosome.

Patch recombinant DNA results from a Holliday junction being resolved by cutting the exchanged strands. The duplex is largely unchanged, except for a DNA sequence on one strand that came from the homologous chromosome.

Key Concepts

- The key event in recombination between two duplex DNA molecules is exchange of single strands.
- When a single strand from one duplex displaces its counterpart in the other duplex, it creates a branched structure.
- The exchange generates a stretch of heteroduplex DNA consisting of one strand from each parent.
- Two (reciprocal) exchanges are necessary to generate a joint molecule.
- The joint molecule is resolved into two separate duplex molecules by nicking two of the connecting strands.
- Whether recombinants are formed depends on whether the strands involved in the

original exchange or the other pair of strands are nicked during resolution.

The act of connecting two duplex molecules of DNA is at the heart of the recombination process. Our molecular analysis of recombination therefore starts by expanding our view of the use of base pairing between complementary single strands in recombination. It is useful to imagine the recombination reaction in terms of single-strand exchanges (although we shall see that this is not necessarily how it is actually initiated), because the properties of the molecules created in this way are central to understanding the processes involved in recombination.

Figure 15.5 illustrates a process that starts with breakage at the corresponding points of the homologous strands of two paired DNA duplexes. The breakage allows movement of the free ends created by the nicks. Each strand leaves its partner and crosses over to pair with its complement in the other duplex.

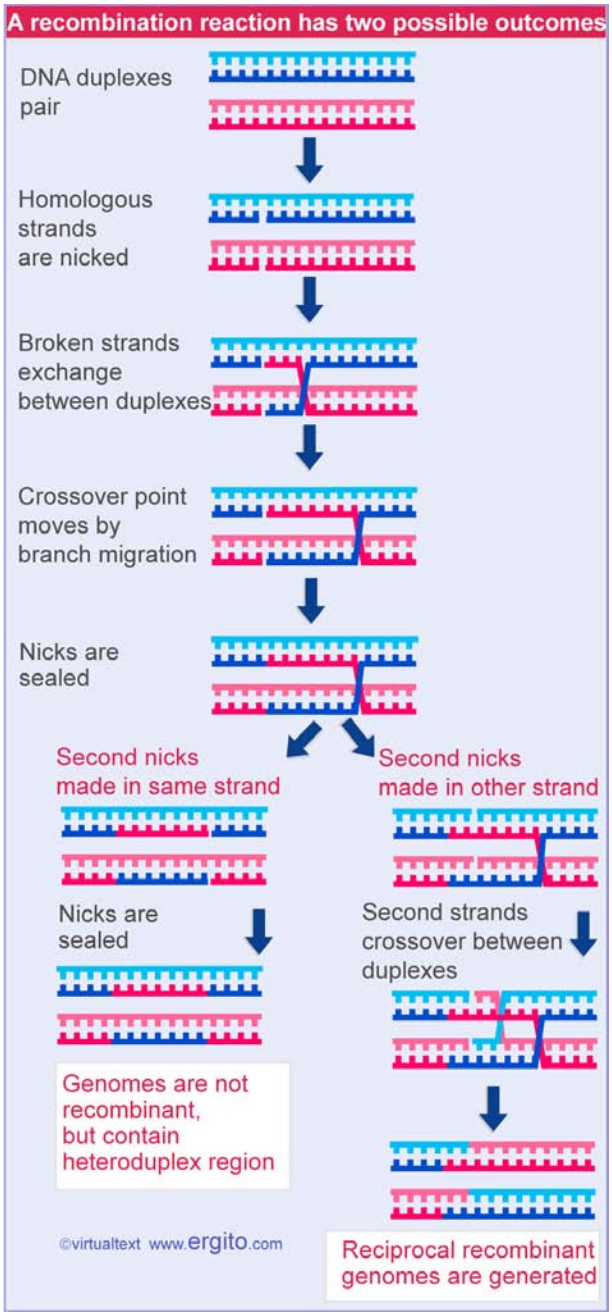

Figure 15.5 Recombination between two paired duplex DNAs could involve reciprocal single-strand exchange, branch migration, and nicking.

The reciprocal exchange creates a connection between the two DNA duplexes. The connected pair of duplexes is called a **joint molecule**. The point at which an individual strand of DNA crosses from one duplex to the other is called the **recombinant joint**.

At the site of recombination, each duplex has a region consisting of one strand from each of the parental DNA molecules. This region is called **hybrid DNA** or

heteroduplex DNA.

An important feature of a recombinant joint is its ability to move along the duplex. Such mobility is called **branch migration**. **Figure 15.6** illustrates the migration of a single strand in a duplex. The branching point can migrate in either direction as one strand is displaced by the other.

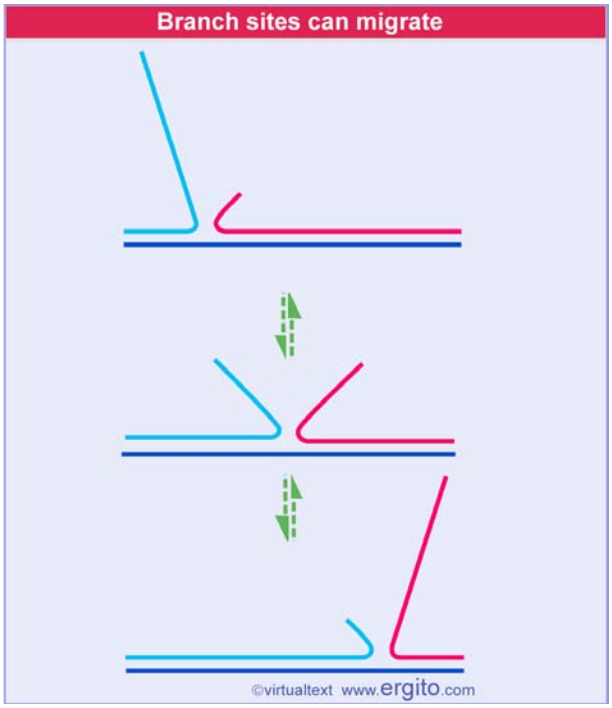

Figure 15.6 Branch migration can occur in either direction when an unpaired single strand displaces a paired strand.

Branch migration is important for both theoretical and practical reasons. As a matter of principle, it confers a dynamic property on recombining structures. As a practical feature, its existence means that the point of branching cannot be established by examining a molecule *in vitro* (because the branch may have migrated since the molecule was isolated).

Branch migration could allow the point of crossover in the recombination intermediate to move in either direction. The rate of branch migration is uncertain, but as seen *in vitro* is probably inadequate to support the formation of extensive regions of heteroduplex DNA in natural conditions. Any extensive branch migration *in vivo* must therefore be catalyzed by a recombination enzyme.

The joint molecule formed by strand exchange must be *resolved* into two separate duplex molecules. **Resolution** requires a further pair of nicks. We can most easily visualize the outcome by viewing the joint molecule in one plane as a **Holliday** structure. This is illustrated in **Figure 15.7**, which represents the structure of **Figure 15.5** with one duplex rotated relative to the other. The outcome of the reaction depends on which pair of strands is nicked.

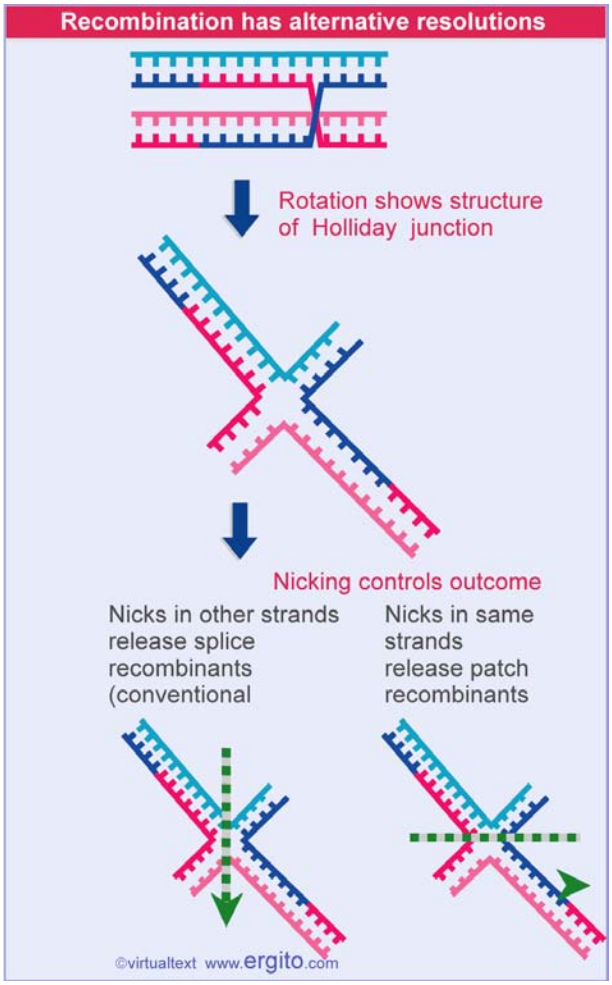

Figure 15.7 Resolution of a Holliday junction can generate parental or recombinant duplexes, depending on which strands are nicked. Both types of product have a region of heteroduplex DNA.

If the nicks are made in the pair of strands that were not originally nicked (the pair that did not initiate the strand exchange), all four of the original strands have been nicked. This releases **splice recombinant** DNA molecules. The duplex of one DNA parent is covalently linked to the duplex of the other DNA parent, via a stretch of heteroduplex DNA. There has been a conventional recombination event between markers located on either side of the heteroduplex region.

If the same two strands involved in the original nicking are nicked again, the other two strands remain intact. The nicking releases the original parental duplexes, which remain intact except that each has a residuum of the event in the form of a length of heteroduplex DNA. These are called **patch recombinants**.

These alternative resolutions of the joint molecule establish the principle that *a strand exchange between duplex DNAs always leaves behind a region of heteroduplex DNA, but the exchange may or may not be accompanied by recombination of the flanking regions.*

What is the minimum length of the region required to establish the connection between the recombining duplexes? Experiments in which short homologous sequences carried by plasmids or phages are introduced into bacteria suggest that the rate of recombination is substantially reduced if the homologous region is <75 bp. This distance is appreciably longer than the ~10 bp required for association between complementary single-stranded regions, which suggests that recombination imposes demands beyond annealing of complements as such.

RECOMBINATION AND REPAIR

4.15.4 Double-strand breaks initiate recombination

Key Terms

A **double-strand break (DSB)** occurs when both strands of a DNA duplex are cleaved at the same site. Genetic recombination is initiated by double-strand breaks. The cell also has repair systems that act on double-strand breaks created at other times.

Key Concepts

- Recombination is initiated by making a double-strand break in one (recipient) DNA duplex.
- Exonuclease action generates 3′–single-stranded ends that invade the other (donor) duplex.
- New DNA synthesis replaces the material that has been degraded.
- This generates a recombinant joint molecule in which the two DNA duplexes are connected by heteroduplex DNA.

The general model of **Figure 15.4** shows that a break must be made in one duplex in order to generate a point from which single strands can unwind to participate in genetic exchange. Both strands of a duplex must be broken to accomplish a genetic exchange. **Figure 15.5** shows a model in which individual breaks in single strands occur successively. However, genetic exchange is actually initiated by a **double-strand break** (DSB). The model is illustrated in **Figure 15.8** (3120; for review see 3121).

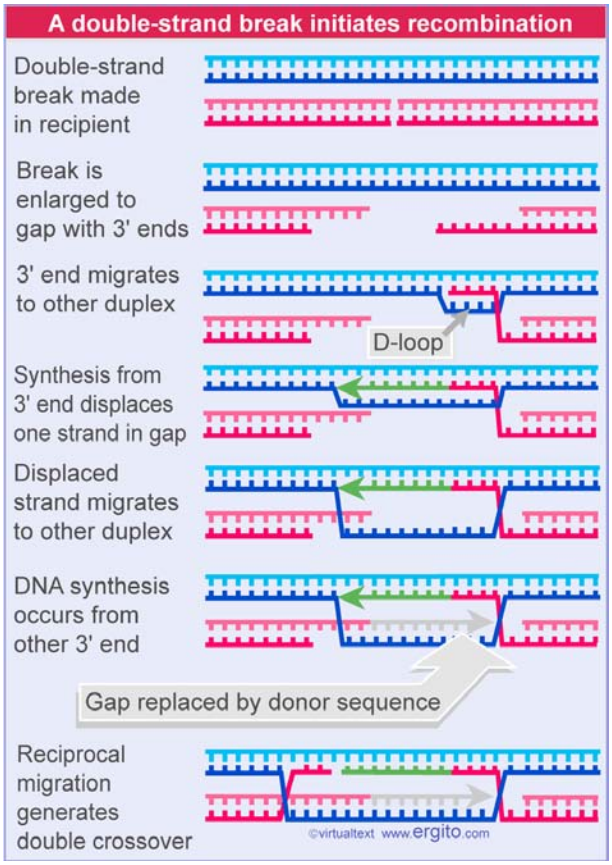

Figure 15.8 Recombination is initiated by a double-strand break, followed by formation of single-stranded 3′ ends, one of which migrates to a homologous duplex.

Recombination is initiated by an endonuclease that cleaves one of the partner DNA duplexes, the "recipient." The cut is enlarged to a gap by exonuclease action. The exonuclease(s) nibble away one strand on either side of the break, generating 3′ single-stranded termini. One of the free 3′ ends then invades a homologous region in the other, "donor" duplex. This is called single-strand invasion (1957). The formation of heteroduplex DNA generates a D loop, in which one strand of the donor duplex is displaced. The D loop is extended by repair DNA synthesis, using the free 3′ end as a primer to generate double-stranded DNA.

Eventually the D loop becomes large enough to correspond to the entire length of the gap on the recipient chromatid. When the extruded single strand reaches the far side of the gap, the complementary single-stranded sequences anneal. Now there is heteroduplex DNA on either side of the gap, and the gap itself is represented by the single-stranded D loop.

The duplex integrity of the gapped region can be restored by repair synthesis using the 3′ end on the left side of the gap as a primer. Overall, the gap has been repaired by two individual rounds of single-strand DNA synthesis.

Branch migration converts this structure into a molecule with two recombinant joints. The joints must be resolved by cutting.

If both joints are resolved in the same way, the original noncrossover molecules will be released, each with a region of altered genetic information that is a footprint of the exchange event. If the two joints are resolved in opposite ways a genetic crossover is produced.

The structure of the two-jointed molecule before it is resolved illustrates a critical difference between the double-strand break model and models that invoke only single-strand exchanges.

- Following the double-strand break, heteroduplex DNA has been formed at each end of the region involved in the exchange. Between the two heteroduplex segments is the region corresponding to the gap, which now has the sequence of the donor DNA in both molecules (**Figure 15.8**). So the arrangement of heteroduplex sequences is asymmetric, and part of one molecule has been converted to the sequence of the other (which is why the initiating chromatid is called the recipient).
- Following reciprocal single-strand exchange, each DNA duplex has heteroduplex material covering the region from the initial site of exchange to the migrating branch (**Figure 15.5**). In variants of the single-strand exchange model in which some DNA is degraded and resynthesized, the initiating chromatid is the donor of genetic information.

The double-strand break model does not reduce the importance of the formation of heteroduplex DNA, which remains the only plausible means by which two duplex molecules can interact. However, by shifting the responsibility for initiating recombination from single-strand to double-strand breaks, it influences our perspective about the ability of the cell to manipulate DNA.

The involvement of double-strand breaks seems surprising at first sight. Once a break has been made right across a DNA molecule, there is no going back. Compare the events of **Figure 15.5** and **Figure 15.8**. In the single-strand exchange model, at no point has any information been lost. But in the double-strand break model, the initial cleavage is immediately followed by loss of information. Any error in retrieving the information could be fatal. On the other hand, the very ability to retrieve lost information by resynthesizing it from another duplex provides a major safety net for the cell.

Last updated on 8-23-2001

Reviews

3120. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983). *The double-strand-break repair model for recombination.* Cell 33, 25-35.

3121. Lichten, M. and Goldman, A. S. (1995). *Meiotic recombination hotspots.* Annu. Rev. Genet. 29, 423-444.

References

1957. Hunter, N. and Kleckner, N. (2001). *The single-end invasion: an asymmetric intermediate at the double-strand break to double-Holliday junction transition of meiotic recombination.* Cell 106, 59-70.

RECOMBINATION AND REPAIR

4.15.5 Recombining chromosomes are connected by the synaptonemal complex

Key Terms

The **synaptonemal complex** describes the morphological structure of synapsed chromosomes.

An **axial element** is a proteinaceous structure around which the chromosomes condense at the start of synapsis.

A **lateral element** is a structure in the synaptonemal complex. It is an axial element that is aligned with the axial elements of other chromosomes.

The **central element** is a structure that lies in the middle of the synaptonemal complex, along which the lateral elements of homologous chromosomes align.

Recombination nodules (Node) are dense objects present on the synaptonemal complex; they may represent protein complexes involved in crossing-over.

Key Concepts

- During the early part of meiosis, homologous chromosomes are paired in the synaptonemal complex.
- The mass of chromatin of each homologue is separated from the other by a proteinaceous complex.

A basic paradox in recombination is that the parental chromosomes never seem to be in close enough contact for recombination of DNA to occur. The chromosomes enter meiosis in the form of replicated (sister chromatid) pairs, visible as a mass of chromatin. They pair to form the **synaptonemal complex**, and it has been assumed for many years that this represents some stage involved with recombination, possibly a necessary preliminary to exchange of DNA. A more recent view is that the synaptonemal complex is a consequence rather than a cause of recombination (for review see 3123). However, we have yet to define how the structure of the synaptonemal complex relates to molecular contacts between DNA molecules (for review see 985).

Synapsis begins when each chromosome (sister chromatid pair) condenses around a structure called the **axial element**, which is apparently proteinaceous. Then the axial elements of corresponding chromosomes become aligned, and the synaptonemal complex forms as a tripartite structure, in which the axial elements, now called **lateral elements**, are separated from each other by a **central element**. **Figure 15.9** shows an example.

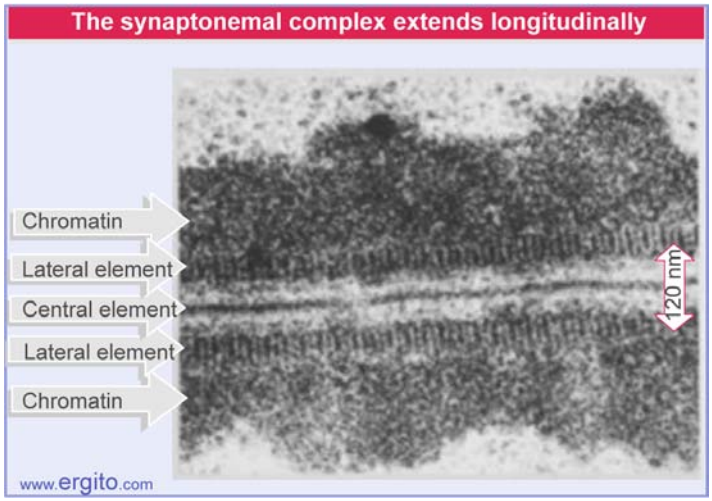

Figure 15.9 The synaptonemal complex brings chromosomes into juxtaposition. This example of *Neotellia* was kindly provided by M. Westergaard and D. Von Wettstein.

Each chromosome at this stage appears as a mass of chromatin bounded by a lateral element. The two lateral elements are separated from each other by a fine but dense central element. The triplet of parallel dense strands lies in a single plane that curves and twists along its axis. The distance between the homologous chromosomes is considerable in molecular terms, more than 200 nm (the diameter of DNA is 2 nm). So a major problem in understanding the role of the complex is that, although it aligns homologous chromosomes, it is far from bringing homologous DNA molecules into contact.

The only visible link between the two sides of the synaptonemal complex is provided by spherical or cylindrical structures observed in fungi and insects. They lie across the complex and are called **nodes** or **recombination nodules**; they occur with the same frequency and distribution as the chiasmata. Their name reflects the hope that they may prove to be the sites of recombination.

From mutations that affect synaptonemal complex formation, we can relate the types of proteins that are involved to its structure. **Figure 15.10** presents a molecular view of the synaptonemal complex. Its distinctive structural features are due to two groups of proteins:

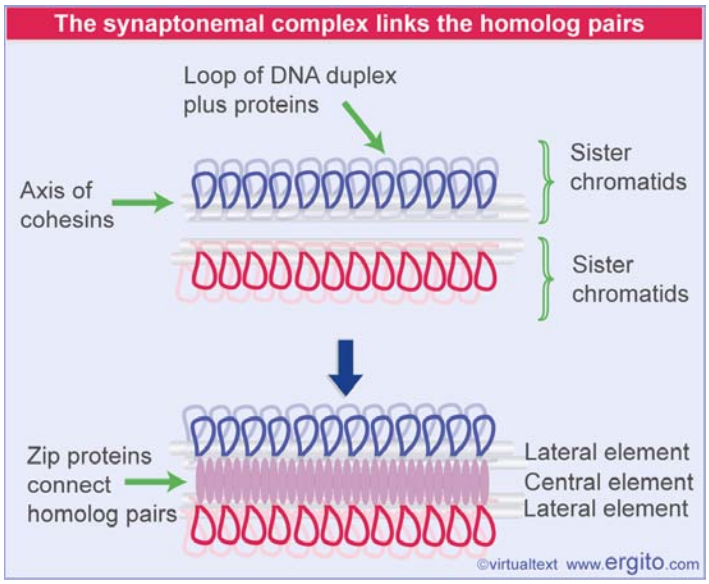

Figure 15.10 Each pair of sister chromatids has an axis made of cohesins. Loops of chromatin project from the axis. The synaptonemal complex is formed by linking together the axes via zip proteins.

- The cohesins form a single linear axis for each pair of sister chromatids from which loops of chromatin extend. This is equivalent to the lateral element of **Figure 15.9**. (The cohesins belong to a general group of proteins involved in connecting sister chromatids so they segregate properly at mitosis of meiosis; see *Molecular Biology 6.29.19 Cohesins hold sister chromatids together*).

- The lateral elements are connected by transverse filaments that are equivalent to the central element of **Figure 15.9**. These are formed from Zip proteins.

Mutations in proteins that are needed for lateral elements to form are found in the genes coding for cohesins. The cohesins that are used in meiosis include Smc3p (which is also used in mitosis) and Rec8p (which is specific to meiosis and is related to the mitotic cohesin Scc1p) (937). The cohesins appear to bind to specific sites along the chromosomes in both mitosis and meiosis (938). They are likely to play a structural role in chromosome segregation. At meiosis, the formation of the lateral elements may be necessary for the later stages of recombination, because although these mutations do not prevent the formation of double-strand breaks, they do block formation of recombinants.

The *zip1* mutation allows lateral elements to form and to become aligned, but they do not become closely synapsed (3136). The N-terminal domain of Zip1 protein is localized in the central element, but the C-terminal domain is localized in the lateral elements (3137). Two other proteins, Zip2 and Zip3 are also localized with Zip1. The group of Zip proteins form transverse filaments that connect the lateral elements of the sister chromatid pairs.

Last updated on 1-3-2003

Reviews

985. Zickler, D. and Kleckner, N. (1999). *Meiotic chromosomes: integrating structure and function.* Annu. Rev. Genet. 33, 603-754.

3123. Roeder, G. S. (1997). *Meiotic chromosomes: it takes two to tango.* Genes Dev. 11, 2600-2621.

References

937. Klein, F. et al. (1999). *A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis.* Cell 98, 91-103.

938. Blat, Y, and Kleckner, N. (1999). *Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the central region.* Cell 98, 249-259.

3136. Sym, M., Engebrecht, J. A., and Roeder, G. S. (1993). *ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis.* Cell 72, 365-378.

3137. Dong, H. and Roeder, G. S. (2000). *Organization of the yeast Zip1 protein within the central region of the synaptonemal complex.* J. Cell Biol. 148, 417-426.

RECOMBINATION AND REPAIR

4.15.6 The synaptonemal complex forms after double-strand breaks

Key Concepts

- Double-strand breaks that initiate recombination occur before the synaptonemal complex forms.
- If recombination is blocked, the synaptonemal complex cannot form.

There is good evidence in yeast that double strand breaks initiate recombination in both homologous and site-specific recombination. Double-strand breaks were initially implicated in the change of mating type, which involves the replacement of one sequence by another (see *Molecular Biology 4.18.8 Unidirectional transposition is initiated by the recipient MAT locus*). Double-strand breaks also occur early in meiosis at sites that provide hotspots for recombination. Their locations are not sequence-specific. They tend to occur in promoter regions and in general to coincide with more accessible regions of chromatin. The frequency of recombination declines in a gradient on one or both sides of the hotspot. The hotspot identifies the site at which recombination is initiated; and the gradient reflects the probability that the recombination events will spread from it (for review see 3122).

We may now interpret the role of double-strand breaks in molecular terms. The flush ends created by the double-strand break are rapidly converted on both sides into long 3′ single-stranded ends, as shown in the model of **Figure 15.8**. A yeast mutation (*rad50*) that blocks the conversion of the flush end into the single-stranded protrusion is defective in recombination. This suggests that double-strand breaks are necessary for recombination. The gradient is determined by the declining probability that a single-stranded region will be generated as distance increases from the site of the double-strand break.

In *rad50* mutants, the 5′ ends of the double-strand breaks are connected to the protein Spo11, which is homologous to the catalytic subunits of a family of type II topoisomerases. This suggests that Spo11 may be a topoisomerase-like enzyme that generates the double-strand breaks. The model for this reaction shown in **Figure 15.11** suggests that Spo11 interacts reversibly with DNA; the break is converted into a permanent structure by an interaction with another protein that dissociates the Spo11 complex. Then removal of Spo11 is followed by nuclease action. At least 9 other proteins are required to process the double-strand breaks. One group of proteins is required to convert the double-strand breaks into protruding 3′ –OH single-stranded ends. Another group then enables the single-stranded ends to invade homologous duplex DNA

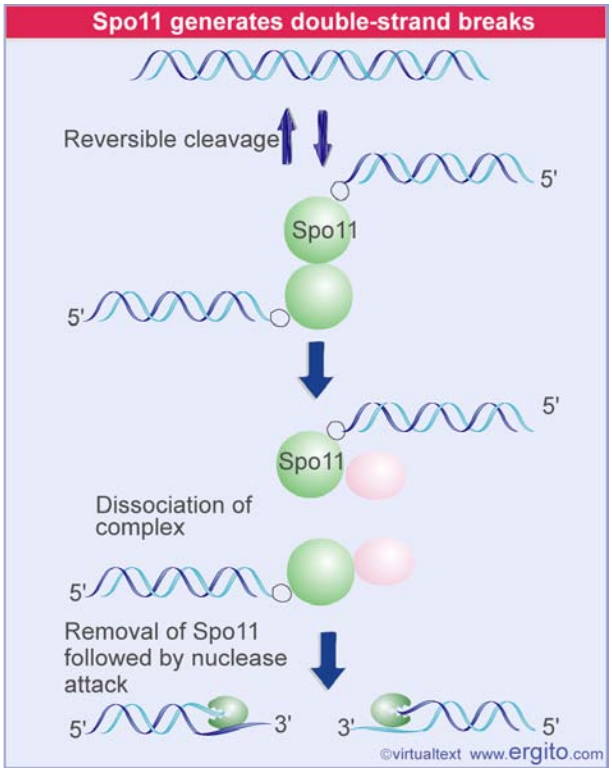

Figure 15.11 Spo11 is covalently joined to the 5′ ends of double-strand breaks.

The correlation between recombination and synaptonemal complex formation is well established, and recent work has shown that all mutations that abolish chromosome pairing in *Drosophila* or in yeast also prevent recombination (for review see 3236). The system for generating the double-strand breaks that initiate recombination is generally conserved. Spo11 homologues have been identified in several higher eukaryotes, and a mutation in the *Drosophila* gene blocks all meiotic recombination.

There are few systems in which it is possible to compare molecular and cytological events at recombination, but recently there has been progress in analyzing meiosis in *S. cerevisiae*. The relative timing of events is summarized in **Figure 15.12**.

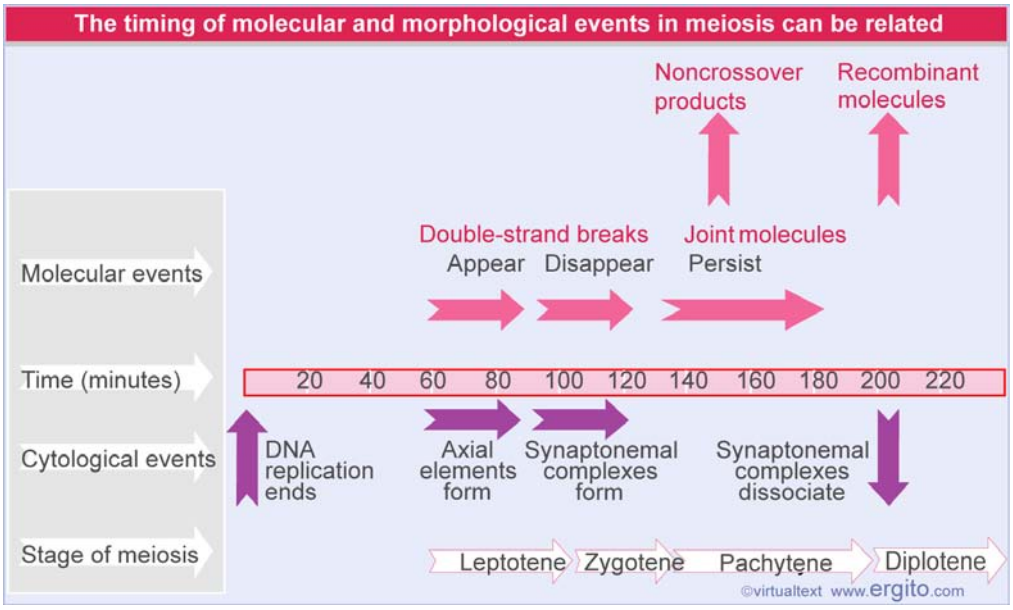

Figure 15.12 Double-strand breaks appear when axial elements form, and disappear during the extension of synaptonemal complexes. Joint molecules appear and persist until DNA recombinants are detected at the end of pachytene.

Double-strand breaks appear and then disappear over a 60 minute period. The first joint molecules, which are putative recombination intermediates, appear soon after the double-strand breaks disappear. The sequence of events suggests that double-strand breaks, individual pairing reactions, and formation of recombinant structures occur in succession at the same chromosomal site (3124).

Double-strand breaks appear during the period when axial elements form. They disappear during the conversion of the paired chromosomes into synaptonemal complexes. This relative timing of events suggests that formation of the synaptonemal complex results from the initiation of recombination via the introduction of double-strand breaks and their conversion into later intermediates of recombination. This idea is supported by the observation that the *rad50* mutant cannot convert axial elements into synaptonemal complexes. This refutes the traditional view of meiosis that the synaptonemal complex represents the need for chromosome pairing to precede the molecular events of recombination.

It has been difficult to determine whether recombination occurs at the stage of synapsis, because recombination is assessed by the appearance of recombinants after the completion of meiosis. However, by assessing the appearance of recombinants in yeast directly in terms of the production of DNA molecules containing diagnostic restriction sites, it has been possible to show that recombinants appear at the end of pachytene. This clearly places the completion of the recombination event after the formation of synaptonemal complexes.

So the synaptonemal complex forms after the double-strand breaks that initiate recombination, and it persists until the formation of recombinant molecules. It does not appear to be necessary for recombination as such, because some mutants that lack a normal synaptonemal complex can generate recombinants. Mutations that abolish recombination, however, also fail to develop a synaptonemal complex. This

suggests that the synaptonemal complex forms as a consequence of recombination, following chromosome pairing, and is required for later stages of meiosis.

Ever since the model for recombination via a Holliday structure was proposed, it has been assumed that the resolution of this structure gives rise to either noncrossover products (with a residual stretch of hybrid DNA) or to crossovers (recombinants), depending on which strands are involved in resolution (see **Figure 15.7**). However, recent measurements of the times of production of noncrossover and crossover molecules suggest that this may not be true (1956). Crossovers do not appear until well after the first appearance of joint molecules, but noncrossovers appear almost simultaneously with the joint molecules (see **Figure 15.12**). If both types of product were produced by the same resolution process, however, we would expect them to appear at the same time. The discrepancy in timing suggests that crossovers are produced as previously thought, by resolution of joint molecules, but that there may be some other route for the production of noncrossovers.

Last updated on 12-2-2002

Reviews

3122. Petes, T. D. (2001). *Meiotic recombination hot spots and cold spots.* Nat. Rev. Genet. 2, 360-369.

3236. McKim, K. S., Jang, J. K., and Manheim, E. A. (2002). *Meiotic recombination and chromosome segregation in Drosophila females.* Annu. Rev. Genet. 36, 205-232.

References

1956. Allers, T. and Lichten, M. (2001). *Differential timing and control of noncrossover and crossover recombination during meiosis.* Cell 106, 47-57.

3124. Weiner, B. M. and Kleckner, N. (1994). *Chromosome pairing via multiple interstitial interactions before and during meiosis in yeast.* Cell 77, 977-991.

RECOMBINATION AND REPAIR

4.15.7 Pairing and synaptonemal complex formation are independent

Key Terms

Crossover control limits the number of recombination events between meiotic chromosomes to 1-2 crossovers per pair of homologs.

Key Concepts

- Mutations can occur in either chromosome pairing or synaptonemal complex formation without affecting the other process.

We can distinguish the processes of pairing and synaptonemal complex formation by the effects of two mutations, each of which blocks one of the processes without affecting the other.

The *zip2* mutation allows chromosomes to pair, but they do not form synaptonemal complexes. So recognition between homologues is independent of recombination or synaptonemal complex formation.

The specificity of association between homologous chromosomes is controlled by the gene *hop2* in *S. cerevisiae*. In *hop2* mutants, normal amounts of synaptonemal complex form at meiosis, but the individual complexes contain nonhomologous chromosomes. This suggests that the formation of synaptonemal complexes as such is independent of homology (and therefore cannot be based on any extensive comparison of DNA sequences). The usual role of Hop2 is to prevent nonhomologous chromosomes from interacting.

Double-strand breaks form in the mispaired chromosomes in the synaptonemal complexes of *hop2* mutants, but they are not repaired. This suggests that, if formation of the synaptonemal complex requires double-strand breaks, it does not require any extensive reaction of these breaks with homologous DNA.

It is not clear what usually happens during pachytene, before DNA recombinants are observed. It may be that this period is occupied by the subsequent steps of recombination, involving the extension of strand exchange, DNA synthesis, and resolution.

At the next stage of meiosis (diplotene), the chromosomes shed the synaptonemal complex; then the chiasmata become visible as points at which the chromosomes are connected. This has been presumed to indicate the occurrence of a genetic exchange, but the molecular nature of a chiasma is unknown. It is possible that it represents the residuum of a completed exchange, or that it represents a connection between homologous chromosomes where a genetic exchange has not yet been resolved. Later in meiosis, the chiasmata move toward the ends of the chromosomes. This flexibility suggests that they represent some remnant of the recombination event,

rather than providing the actual intermediate.

Recombination events occur at discrete points on meiotic chromosomes, but we cannot as yet correlate their occurrence with the discrete structures that have been observed, that is, recombination nodules and chiasmata. However, insights into the molecular basis for the formation of discontinuous structures are provided by the identification of proteins involved in yeast recombination that can be localized to discrete sites. These include MSH4 (which is related to bacterial proteins involved in mismatch-repair), and Dmc1 and Rad51 (which are homologues of the *E. coli* RecA protein). The exact roles of these proteins in recombination remain to be established.

Recombination events are subject to a general control. Only a minority of interactions actually mature as crossovers, but these are distributed in such a way that typically each pair of homologues acquires only 1-2 crossovers, yet the probability of zero crossovers for a homologue pair is very low (<0.1%). This process is probably the result of a single **crossover control**, because the nonrandomness of crossovers is generally disrupted in certain mutants. Furthermore, the occurrence of recombination is necessary for progress through meiosis, and a "checkpoint" system (see *Molecular Biology 6.29 Cell cycle and growth regulation*) exists to block meiosis if recombination has not occurred. (The block is lifted when recombination has been successfully completed; this system provides a safeguard to ensure that cells do not try to segregate their chromosomes until recombination has occurred.)

Last updated on 1-21-2000

RECOMBINATION AND REPAIR

4.15.8 The bacterial RecBCD system is stimulated by *chi* sequences

Key Concepts

- The RecBCD complex has nuclease and helicases activities.
- It binds to DNA downstream of a *chi* sequence, unwinds the duplex, and degrades one strand from 3′–5′ as it moves to the *chi* site.
- The *chi* site triggers loss of the RecD subunit and nuclease activity.

The nature of the events involved in exchange of sequences between DNA molecules was first described in bacterial systems. Here the recognition reaction is part and parcel of the recombination mechanism and involves restricted regions of DNA molecules rather than intact chromosomes. But the general order of molecular events is similar: a single strand from a broken molecule interacts with a partner duplex; the region of pairing is extended; and an endonuclease resolves the partner duplexes. Enzymes involved in each stage are known, although they probably represent only some of the components required for recombination.

Bacterial enzymes implicated in recombination have been identified by the occurrence of ***rec***$^-$ mutations in their genes. The phenotype of Rec$^-$ mutants is the inability to undertake generalized recombination. Some 10-20 loci have been identified.

Bacteria do not usually exchange large amounts of duplex DNA, but there may be various routes to initiate recombination in prokaryotes. In some cases, DNA may be available with free single-stranded 3′ ends: DNA may be provided in single-stranded form (as in conjugation; see *Molecular Biology 4.13.13 Conjugation transfers single-stranded DNA*); single-stranded gaps may be generated by irradiation damage; or single-stranded tails may be generated by phage genomes undergoing replication by a rolling circle. However, in circumstances involving two duplex molecules (as in recombination at meiosis in eukaryotes), single-stranded regions and 3′ ends must be generated.

One mechanism for generating suitable ends has been discovered as a result of the existence of certain hotspots that stimulate recombination. They were discovered in phage lambda in the form of mutants, called *chi*, that have single base-pair changes creating sequences that stimulate recombination. These sites lead us to the role of other proteins involved in recombination.

These sites share a constant nonsymmetrical sequence of 8 bp:

5′ GCTGGTGG 3′
3′ CGACCACC 5′

The *chi* sequence occurs naturally in *E. coli* DNA about once every 5-10 kb. Its absence from wild-type lambda DNA, and also from other genetic elements, shows that it is not essential for recombination.

A *chi* sequence stimulates recombination in its general vicinity, say within a distance of up to 10 kb from the site. A *chi* site can be activated by a double-strand break made several kb away on one particular side (to the right of the sequence as written above). This dependence on orientation suggests that the recombination apparatus must associate with DNA at a broken end, and then can move along the duplex only in one direction.

chi sites are targets for the action of an enzyme coded by the genes *recBCD*. This complex exercises several activities. It is a potent nuclease that degrades DNA, originally identified as the activity exonuclease V. It has helicase activities that can unwind duplex DNA in the presence of SSB; and it has an ATPase activity. Its role in recombination may be to provide a single-stranded region with a free 3′ end.

Figure 15.13 shows how these reactions are coordinated on a substrate DNA that has a *chi* site. RecBCD binds to DNA at a double-stranded end. Two of its subunits have helicase activities: RecD functions with 5′-3′ polarity; and RecB functions with 3′-5′ polarity (4182, 4183). Translocation along DNA and unwinding the double helix is initially driven by the RecD subunit. As RecBCD advances, it degrades the released single strand with the 3′ end. When it reaches the *chi* site, it recognizes the top strand of the *chi* site in single-stranded form. This causes the enzyme to pause. It then cleaves the top strand of the DNA at a position between 4 and 6 bases to the right of *chi*. Recognition of the *chi* site causes the RecD subunit to dissociate or become inactivated, as a result of which the enzyme loses its nuclease activity. However, it continues to function as a helicase, now using only the RecB subunit to drive translocation, at about half the previous speed (4184). The overall result of this interaction is to generate single-stranded DNA with a 3′ end at the *chi* sequence. This is a substrate for recombination.

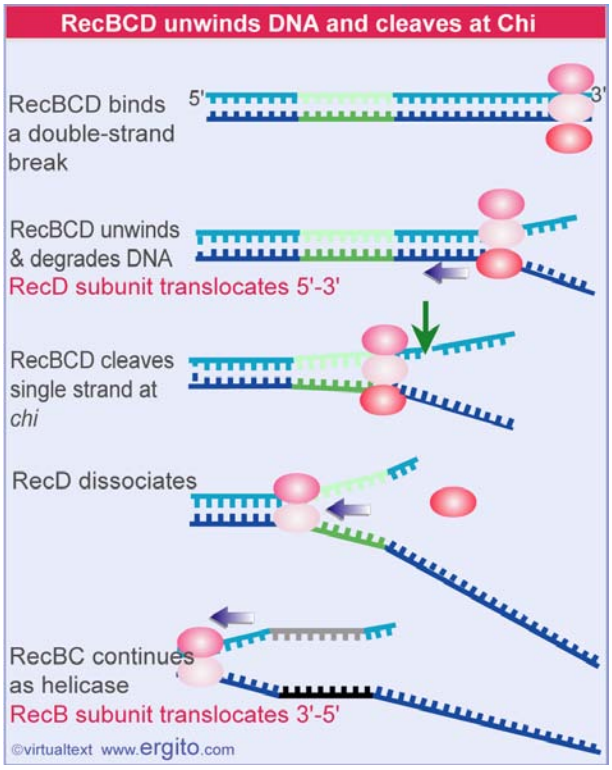

Figure 15.13 RecBCD nuclease approaches a *chi* sequence from one side, degrading DNA as it proceeds; at the *chi* site, it makes an endonucleolytic cut, loses RecD, and retains only the helicase activity.

Last updated on 9-26-2003

References

4182. Taylor, A. F. and Smith, G. R. (2003). *RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity.* Nature 423, 889-893.

4183. Dillingham, M. S., Spies, M., and Kowalczykowski, S. C. (2003). *RecBCD enzyme is a bipolar DNA helicase.* Nature 423, 893-897.

4184. Spies, M., Bianco, P. R., Dillingham, M. S., Handa, N., Baskin, R. J., and Kowalczykowski, S. C. (2003). *A molecular throttle: the recombination hotspot chi controls DNA translocation by the RecBCD helicase.* Cell 114, 647-654.

RECOMBINATION AND REPAIR

4.15.9 Strand-transfer proteins catalyze single-strand assimilation

Key Terms

Single-strand assimilation (Single-strand uptake) describes the ability of RecA protein to cause a single strand of DNA to displace its homologous strand in a duplex; that is, the single strand is assimilated into the duplex.

A **paranemic joint** describes a region in which two complementary sequences of DNA are associated side by side instead of being intertwined in a double helical structure.

A **plectonemic** joint is a region that consists of one molecule wound around another molecule, e.g. the DNA strands in a double helix.

Key Concepts

- RecA forms filaments with single-stranded or duplex DNA and catalyzes the ability of a single-stranded DNA with a free 3′ to displace its counterpart in a DNA duplex.

The *E. coli* protein RecA was the first example to be discovered of a DNA strand-transfer protein. It is the paradigm for a group that includes several other bacterial and archaeal proteins, Rad51 in *S. cerevisiae*, and the higher eukaryotic protein Dmc1 (for review see 3129; 3128; 3130). Analysis of yeast *rad51* mutants shows that this class of protein plays a central role in recombination. They accumulate double-strand breaks and fail to form normal synaptonemal complexes. This reinforces the idea that exchange of strands between DNA duplexes is involved in formation of the synaptonemal complex, and raises the possibility that chromosome synapsis is related to the bacterial strand assimilation reaction.

RecA in bacteria has two quite different types of activity: it can stimulate protease activity in the SOS response (see *Molecular Biology 4.15.27 RecA triggers the SOS system*); and can promote base pairing between a single strand of DNA and its complement in a duplex molecule. Both activities are activated by single-stranded DNA in the presence of ATP.

The DNA-handling activity of RecA enables a single strand to displace its homologue in a duplex in a reaction that is called **single-strand uptake** or **single-strand assimilation**. The displacement reaction can occur between DNA molecules in several configurations and has three general conditions:

- One of the DNA molecules must have a single-stranded region.
- One of the molecules must have a free 3′ end.
- The single-stranded region and the 3′ end must be located within a region that is

complementary between the molecules.

The reaction is illustrated in **Figure 15.14**. When a linear single strand invades a duplex, it displaces the original partner to its complement. The reaction can be followed most easily by making either the donor or recipient a circular molecule. The reaction proceeds 5′–3′ along the strand whose partner is being displaced and replaced, that is, the reaction involves an exchange in which (at least) one of the exchanging strands has a free 3′ end.

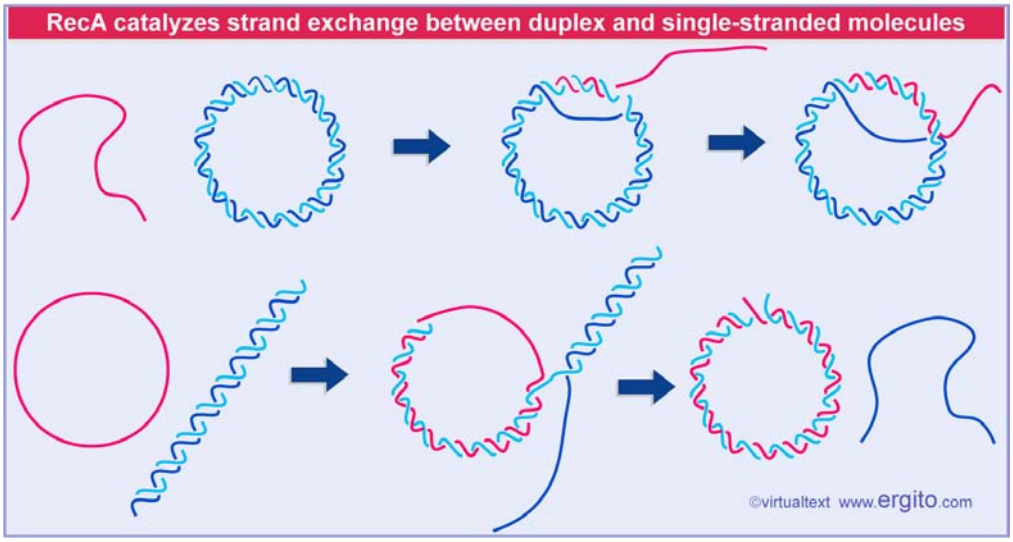

Figure 15.14 RecA promotes the assimilation of invading single strands into duplex DNA so long as one of the reacting strands has a free end.

Single-strand assimilation is potentially related to the initiation of recombination. All models call for an intermediate in which one or both single strands cross over from one duplex to the other (see **Figure 15.5** and **Figure 15.8**). RecA could catalyze this stage of the reaction. In the bacterial context, RecA acts on substrates generated by RecBCD. RecBCD-mediated unwinding and cleavage can be used to generate ends that initiate the formation of heteroduplex joints. RecA can take the single strand with the 3′ end that is released when RecBCD cuts at *chi*, and can use it to react with a homologous duplex sequence, thus creating a joint molecule.

All of the bacterial and archaeal proteins in the RecA family can aggregate into long filaments with single-stranded or duplex DNA. (Eukaryotic homologues of RecA do not form filaments, so the mechanics of the reaction are likely to be different in eukaryotes.) There are 6 RecA monomers per turn of the filament, which has a helical structure with a deep groove that contains the DNA. The stoichiometry of binding is 3 nucleotides (or base pairs) per RecA monomer. The DNA is held in a form that is extended 1.5 times relative to duplex B DNA, making a turn every 18.6 nucleotides (or base pairs). When duplex DNA is bound, it contacts RecA via its minor groove, leaving the major groove accessible for possible reaction with a second DNA molecule.

The interaction between two DNA molecules occurs within these filaments. When a single strand is assimilated into a duplex, the first step is for RecA to bind the single

strand into a filament. Then the duplex is incorporated, probably forming some sort of triple-stranded structure. In this system, synapsis precedes physical exchange of material, because the pairing reaction can take place even in the absence of free ends, when strand exchange is impossible. A free 3′ end is required for strand exchange. The reaction occurs within the filament, and RecA remains bound to the strand that was originally single, so that at the end of the reaction RecA is bound to the duplex molecule.

All of the proteins in this family can promote the basic process of strand exchange without a requirement for energy input. However, RecA augments this activity by using ATP hydrolysis. Large amounts of ATP are hydrolyzed during the reaction. The ATP may act through an allosteric effect on RecA conformation. When bound to ATP, the DNA-binding site of RecA has a high affinity for DNA; this is needed to bind DNA and for the pairing reaction. Hydrolysis of ATP converts the binding site to low affinity, which is needed to release the heteroduplex DNA.

We can divide the reaction that RecA catalyzes between single-stranded and duplex DNA into three phases:

- a slow presynaptic phase in which RecA polymerizes on single-stranded DNA;
- a fast pairing reaction between the single-stranded DNA and its complement in the duplex to produce a heteroduplex joint;
- a slow displacement of one strand from the duplex to produce a long region of heteroduplex DNA.

The presence of SSB (single-strand binding protein) stimulates the reaction, by ensuring that the substrate lacks secondary structure. It is not clear yet how SSB and RecA both can act on the same stretch of DNA. Like SSB, RecA is required in stoichiometric amounts, which suggests that its action in strand assimilation involves binding cooperatively to DNA to form a structure related to the filament.

When a single-stranded molecule reacts with a duplex DNA, the duplex molecule becomes unwound in the region of the recombinant joint. The initial region of heteroduplex DNA may not even lie in the conventional double helical form, but could consist of the two strands associated side by side. A region of this type is called a **paranemic joint** (compared with the classical intertwined **plectonemic** relationship of strands in a double helix). A paranemic joint is unstable; further progress of the reaction requires its conversion to the double-helical form. This reaction is equivalent to removing negative supercoils and may require an enzyme that solves the unwinding/rewinding problem by making transient breaks that allow the strands to rotate about each other.

All of the reactions we have discussed so far represent only a part of the potential recombination event: the invasion of one duplex by a single strand. Two duplex molecules can interact with each other under the sponsorship of RecA, provided that one of them has a single-stranded region of at least 50 bases. The single-stranded region can take the form of a tail on a linear molecule or of a gap in a circular molecule.

The reaction between a partially duplex molecule and an entirely duplex molecule leads to the exchange of strands. An example is illustrated in **Figure 15.15**. Assimilation starts at one end of the linear molecule, where the invading single strand displaces its homologue in the duplex in the customary way. But when the reaction reaches the region that is duplex in both molecules, the invading strand unpairs from its partner, which then pairs with the other displaced strand.

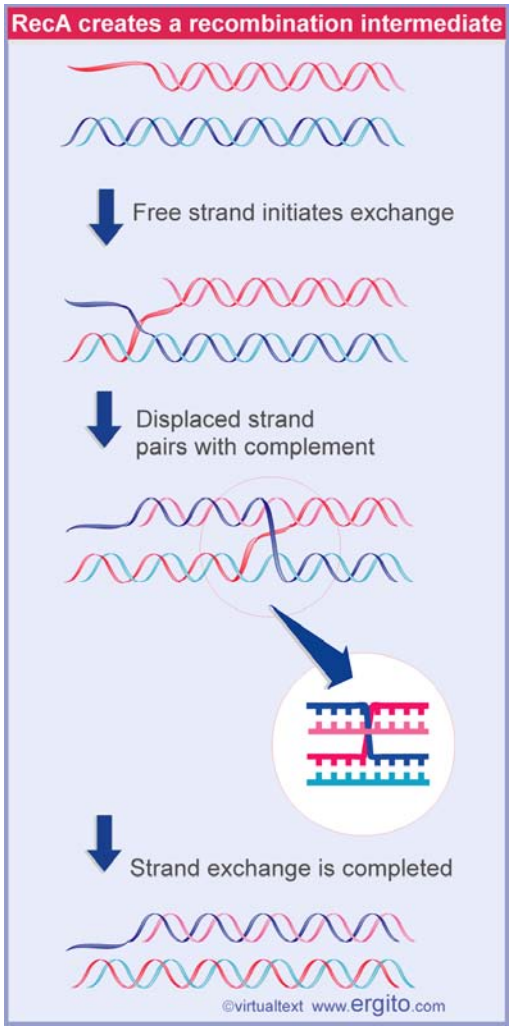

Figure 15.15 RecA-mediated strand exchange between partially duplex and entirely duplex DNA generates a joint molecule with the same structure as a recombination intermediate.

At this stage, the molecule has a structure indistinguishable from the recombinant joint in **Figure 15.7**. The reaction sponsored *in vitro* by RecA can generate Holliday junctions, which suggests that the enzyme can mediate reciprocal strand transfer. We know less about the geometry of four-strand intermediates bound by RecA, but presumably two duplex molecules can lie side by side in a way consistent with the requirements of the exchange reaction.

The biochemical reactions characterized *in vitro* leave open many possibilities for the

functions of strand-transfer proteins *in vivo*. Their involvement is triggered by the availability of a single-stranded 3′ end. In bacteria, this is most likely generated when RecBCD processes a double-strand break to generate a single-stranded end. One of the main circumstances in which this is invoked may be when a replication fork stalls at a site of DNA damage (see *Molecular Biology 4.15.26 Recombination is an important mechanism to recover from replication errors*). The introduction of DNA during conjugation, when RecA is required for recombination with the host chromosome, is more closely related to conventional recombination. In yeast, double-strand breaks may be generated by DNA damage or as part of the normal process of recombination. In either case, processing of the break to generate a 3′ single-stranded end is followed by loading the single strand into a filament with Rad51, followed by a search for matching duplex sequences. This can be used in both repair and recombination reactions.

Last updated on 11-30-2002

Reviews

3128. Kowalczykowski, S. C. and Eggleston, A. K. (1994). *Homologous pairing and DNA strand-exchange proteins.* Annu. Rev. Biochem. 63, 991-1043.

3129. Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., and Rehrauer, W. M. (1994). *Biochemistry of homologous recombination in Escherichia coli.* Microbiol. Rev. 58, 401-465.

3130. Lusetti, S. L. and Cox, M. M. (2002). *The bacterial RecA protein and the recombinational DNA repair of stalled replication forks.* Annu. Rev. Biochem. 71, 71-100.

RECOMBINATION AND REPAIR

4.15.10 The Ruv system resolves Holliday junctions

Key Concepts

- The Ruv complex acts on recombinant junctions.
- RuvA recognizes the structure of the junction and RuvB is a helicase that catalyzes branch migration.
- RuvC cleaves junctions to generate recombination intermediates.

One of the most critical steps in recombination is the resolution of the Holliday junction, which determines whether there is a reciprocal recombination or a reversal of the structure that leaves only a short stretch of hybrid DNA (see **Figure 15.5** and **Figure 15.7**). Branch migration from the exchange site (see **Figure 15.6**) determines the length of the region of hybrid DNA (with or without recombination).The proteins involved in stabilizing and resolving Holliday junctions have been identified as the products of the *Ruv* genes in *E. coli* (for review see 3131). RuvA and RuvB increase the formation of heteroduplex structures. RuvA recognizes the structure of the Holliday junction. RuvA binds to all four strands of DNA at the crossover point and forms two tetramers that sandwich the DNA. RuvB is a hexameric helicase with an ATPase activity that provides the motor for branch migration. Hexameric rings of RuvB bind around each duplex of DNA upstream of the crossover point. A diagram of the complex is shown in **Figure 15.16**.

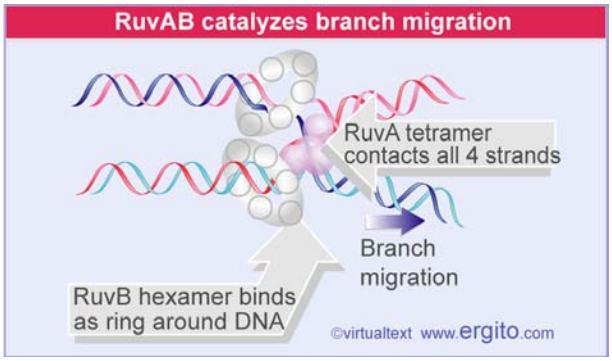

Figure 15.16 RuvAB is an asymmetric complex that promotes branch migration of a Holliday junction.

The RuvAB complex can cause the branch to migrate as fast as 10-20 bp/sec. A similar activity is provided by another helicase, RecG. RuvAB displaces RecA from DNA during its action. The RuvAB and RecG activities both can act on Holliday junctions, but if both are mutant, *E. coli* is completely defective in recombination activity.

The third gene, *ruvC*, codes for an endonuclease that specifically recognizes Holliday junctions. It can cleave the junctions *in vitro* to resolve recombination intermediates.

A common tetranucleotide sequence provides a hotspot for RuvC to resolve the Holliday junction. The tetranucleotide (ATTG) is asymmetric, and thus may direct resolution with regard to which pair of strands is nicked. This determines whether the outcome is patch recombinant formation (no overall recombination) or splice recombinant formation (recombination between flanking markers). Crystal structures of RuvC and other junction-resolving enzymes show that there is a little structural similarity among the group, in spite of their common function (for review see 3449).

All of this suggests that recombination uses a "resolvasome" complex that includes enzymes catalyzing branch migration as well as junction-resolving activity. It is possible that mammalian cells contain a similar complex (3460).

We may now account for the stages of recombination in *E. coli* in terms of individual proteins. **Figure 15.17** shows the events that are involved in using recombination to repair a gap in one duplex by retrieving material from the other duplex. The major caveat in applying these conclusions to recombination in eukaryotes is that bacterial recombination generally involves interaction between a fragment of DNA and a whole chromosome. It occurs as a repair reaction that is stimulated by damage to DNA, and this is not entirely equivalent to recombination between genomes at meiosis. Nonetheless, similar molecular activities are involved in manipulating DNA.

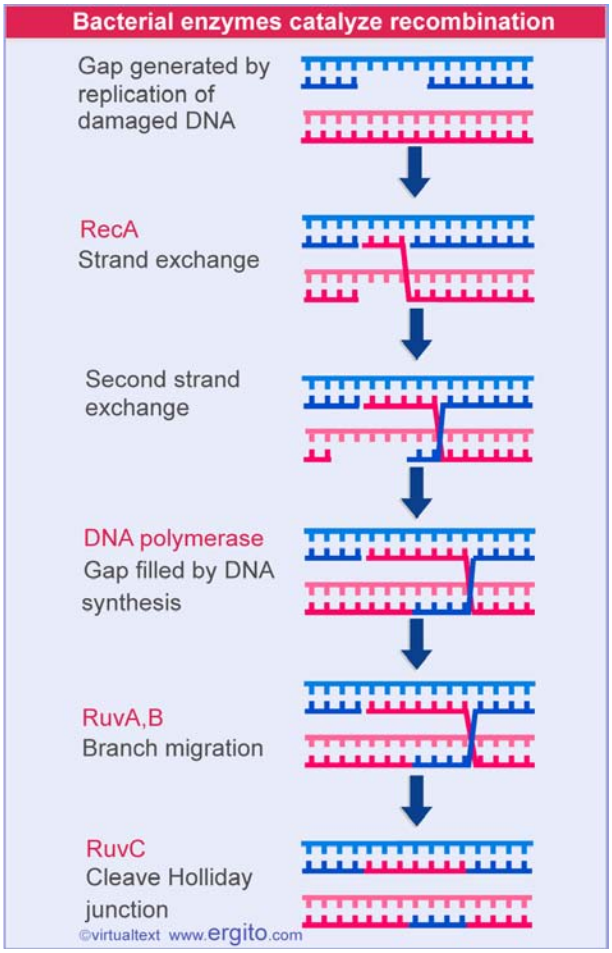

Figure 15.17 Bacterial enzymes can catalyze all stages of recombination in the repair pathway following the production of suitable substrate DNA molecules.

Another system of resolvases has been characterized in yeast and mammals. Mutants in *S. cerevisiae mus81* are defective in recombination. Mus81 is a component of an endonuclease that resolves Holliday junctions into duplex structures (2930; 2931; 2932). The resolvase is important both in meiosis and for restarting stalled replication forks (see *Molecular Biology 4.15.26 Recombination is an important mechanism to recover from replication errors*).

Last updated on 9-16-2002

Reviews

3131. West, S. C. (1997). *Processing of recombination intermediates by the RuvABC proteins.* Annu. Rev. Genet. 31, 213-244.

3449. Lilley, D. M. and White, M. F. (2001). *The junction-resolving enzymes.* Nat. Rev. Mol. Cell Biol. 2, 433-443.

References

2930. Kaliraman, V., Mullen, J. R., Fricke, W. M., Bastin-Shanower, S. A., and Brill, S. J. (2001). *Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease.* Genes Dev. 15, 2730-2740.

2931. Chen, X. B., Melchionna, R., Denis, C. M., Gaillard, P. H., Blasina, A., Van de Weyer, I., Boddy, M. N., Russell, P., Vialard, J., and McGowan, C. H. (2001). *Human Mus81-associated endonuclease cleaves Holliday junctions in vitro.* Mol. Cell 8, 1117-1127.

2932. Boddy, M. N., Gaillard, P. H., McDonald, W. H., Shanahan, P., Yates, J. R., and Russell, P. (2001). *Mus81-Eme1 are essential components of a Holliday junction resolvase.* Cell 107, 537-548.

3460. Constantinou, A., Davies, A. A., and West, S. C. (2001). *Branch migration and Holliday junction resolution catalyzed by activities from mammalian cells.* Cell 104, 259-268.

RECOMBINATION AND REPAIR

4.15.11 Gene conversion accounts for interallelic recombination

Key Terms

The **ascus** of a fungus contains a tetrad or octad of the (haploid) spores, representing the products of a single meiosis.

A **tetrad** describes the four (haploid) spores that result from a meiosis in yeast. (The term originally was used to describe the structure found at the beginning of meiosis, now known as a bivalent, that contains all four chromatids, produced by duplication of a homologous chromosome pair.)

Postmeiotic segregation describes the segregation of two strands of a duplex DNA that bear different information (created by heteroduplex formation during meiosis) when a subsequent replication allows the strands to separate.

Gene conversion is the alteration of one strand of a heteroduplex DNA to make it complementary with the other strand at any position(s) where there were mispaired bases.

Key Concepts

- Heteroduplex DNA that is created by recombination can have mismatched sequences where the recombining alleles are not identical.
- Repair systems may remove mismatches by changing one of the strands so its sequence is complementary to the other.

The involvement of heteroduplex DNA explains the characteristics of recombination between alleles; indeed, allelic recombination provided the impetus for the development of the heteroduplex model. When recombination between alleles was discovered, the natural assumption was that it takes place by the same mechanism of reciprocal recombination that applies to more distant loci. That is to say that an individual breakage and reunion event occurs within the locus to generate a reciprocal pair of recombinant chromosomes. However, in the close quarters of a single gene, the formation of heteroduplex DNA itself is usually responsible for the recombination event.

Individual recombination events can be studied in the Ascomycetes fungi, because the products of a single meiosis are held together in a large cell, the **ascus**, also sometimes called the **tetrad** (see *Genetics 2.7.6 S. cerevisiae meiotic growth and tetrad analysis*). Even better, in some fungi, the four haploid nuclei produced by meiosis are arranged in a linear order. Actually, a mitosis occurs after the production of these four nuclei, giving a linear series of eight haploid nuclei. **Figure 15.18** shows that each of these nuclei effectively represents the genetic character of one of the eight strands of the four chromosomes produced by the meiosis.

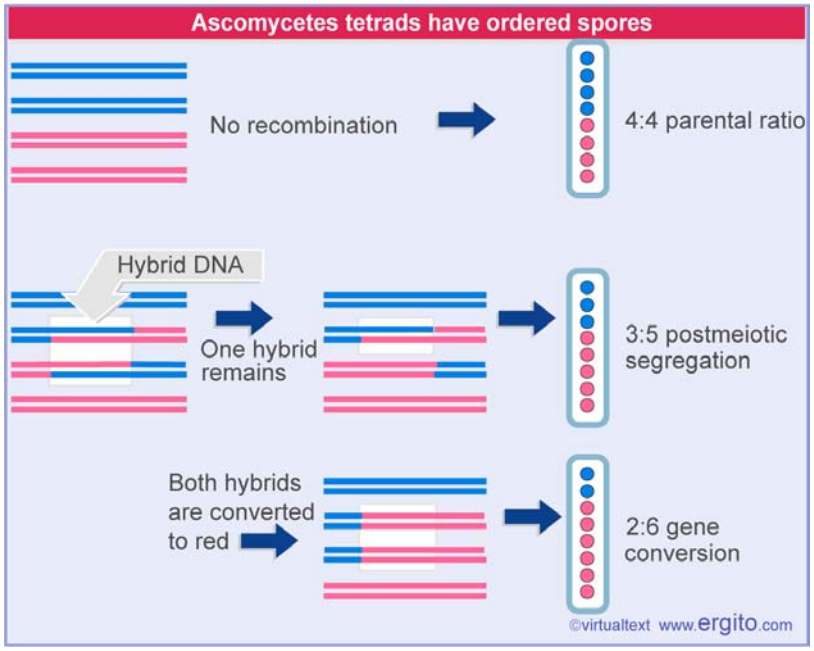

Figure 15.18 Spore formation in the *Ascomycetes* allows determination of the genetic constitution of each of the DNA strands involved in meiosis.

Meiosis in a heterozygote should generate four copies of each allele. This is seen in the majority of spores. But there are some spores with abnormal ratios. They are explained by the formation and correction of heteroduplex DNA in the region in which the alleles differ. The figure illustrates a recombination event in which a length of hybrid DNA occurs on one of the four meiotic chromosomes, a possible outcome of recombination initiated by a double-strand break.

Suppose that two alleles differ by a single point mutation. When a strand exchange occurs to generate heteroduplex DNA, the two strands of the heteroduplex will be mispaired at the site of mutation. So each strand of DNA carries different genetic information. If no change is made in the sequence, the strands separate at the ensuing replication, each giving rise to a duplex that perpetuates its information. This event is called **postmeiotic segregation**, because it reflects the separation of DNA strands after meiosis. Its importance is that it demonstrates directly the existence of heteroduplex DNA in recombining alleles.

Another effect is seen when examining recombination between alleles: the proportions of the alleles differ from the initial 4:4 ratio. This effect is called **gene conversion**. It describes a nonreciprocal transfer of information from one chromatid to another.

Gene conversion results from exchange of strands between DNA molecules, and the change in sequence may have either of two causes at the molecular level:

- As indicated by the double-strand break model in **Figure 15.8**, one DNA duplex may act as a donor of genetic information that directly replaces the corresponding sequences in the recipient duplex by a process of gap generation,

strand exchange, and gap filling.

- As part of the exchange process, heteroduplex DNA is generated when a single strand from one duplex pairs with its complement in the other duplex. Repair systems recognize mispaired bases in heteroduplex DNA, and may then excise and replace one of the strands to restore complementarity. Such an event converts the strand of DNA representing one allele into the sequence of the other allele.

Gene conversion does not depend on crossing-over, but is correlated with it. A large proportion of the aberrant asci show genetic recombination between two markers on either side of a site of interallelic gene conversion. This is exactly what would be predicted if the aberrant ratios result from initiation of the recombination process as shown in **Figure 15.5**, but with an approximately equal probability of resolving the structure with or without recombination (as indicated in **Figure 15.7**). The implication is that fungal chromosomes initiate crossing-over about twice as often as would be expected from the measured frequency of recombination between distant genes.

Various biases are seen when recombination is examined at the molecular level. Either direction of gene conversion may be equally likely, or allele-specific effects may create a preference for one direction. Gradients of recombination may fall away from hotspots. We now know that hotspots represent sites at which double-strand breaks are initiated, and the gradient is correlated with the extent to which the gap at the hotspot is enlarged and converted to long single-stranded ends (see *Molecular Biology 4.15.6 The synaptonemal complex forms after double-strand breaks*).

Some information about the extent of gene conversion is provided by the sequences of members of gene clusters. Usually, the products of a recombination event will separate and become unavailable for analysis at the level of DNA sequence. However, when a chromosome carries two (non-allelic) genes that are related, they may recombine by an "unequal crossing-over" event (see *Molecular Biology 1.4.7 Unequal crossing-over rearranges gene clusters*). All we need to note for now is that a heteroduplex may be formed between the two nonallelic genes. Gene conversion effectively converts one of the nonallelic genes to the sequence of the other.

The presence of more than one gene copy on the same chromosome provides a footprint to trace these events. For example, if heteroduplex formation and gene conversion occurred over part of one gene, this part may have a sequence identical with or very closely related to the other gene, while the remaining part shows more divergence. Available sequences suggest that gene conversion events may extend for considerable distances, up to a few thousand bases.

RECOMBINATION AND REPAIR

4.15.12 Supercoiling affects the structure of DNA

Key Terms

Supercoiling describes the coiling of a closed duplex DNA in space so that it crosses over its own axis.

The **linking number** is the number of times the two strands of a closed DNA duplex cross over each other.

Topological isomers are molecules of DNA that are identical except for a difference in linking number.

The **twisting number** of a DNA is the number of base pairs divided by the number of base pairs per turn of the double helix.

The **writhing number** is the number of times a duplex axis crosses over itself in space.

Key Concepts

- Supercoiling occurs only in a closed DNA with no free ends.
- A closed DNA can be a circular DNA molecule or a linear molecule where both ends are anchored in a protein structure.
- Any closed DNA molecule has a linking number, which is the sum of the twisting number and writhing number.
- Turns can be repartitioned between the twisting number and writhing number, so that a change in the structure of the double helix can compensate for a change in its coiling in space.
- The linking number can be changed only by breaking and making bonds in DNA.

The winding of the two strands of DNA around each other in the double helical structure makes it possible to change the structure by influencing its conformation in space. If the two ends of a DNA molecule are fixed, the double helix can be wound around itself in space. This is called **supercoiling**. The effect can be imagined like a rubber band twisted around itself. The simplest example of a DNA with no fixed ends is a circular molecule. The effect of supercoiling can be seen by comparing the nonsupercoiled circular DNA lying flat in **Figure 15.19** with the supercoiled circular molecule that forms a twisted and therefore more condensed shape.

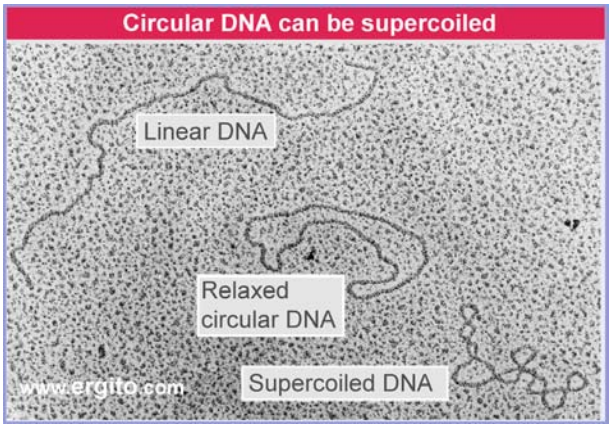

Figure 15.19 Linear DNA is extended, a circular DNA remains extended if it is relaxed (nonsupercoiled), but a supercoiled DNA has a twisted and condensed form.

The consequences of supercoiling depend on whether the DNA is twisted around itself in the same sense as the two strands within the double helix (clockwise) or in the opposite sense. Twisting in the same sense produces *positive supercoiling*. This has the effect of causing the DNA strands to wound around one another more tightly, so that there are more base pairs per turn. Twisting in the opposite sense produces *negative supercoiling*. This causes the DNA strands to be twisted around one another less tightly, so there are fewer base pairs per turn. Negative supercoiling can be thought of as creating tension in the DNA that is relieved by unwinding the double helix. The ultimate effect of negative supercoiling is to generate a region in which the two strands of DNA have separated – formally there are zero base pairs per turn.

Topological manipulation of DNA is a central aspect of all its functional activities – recombination, replication, and transcription – as well as of the organization of higher-order structure. All synthetic activities involving double-stranded DNA require the strands to separate. However, the strands do not simply lie side by side; they are intertwined. Their separation therefore requires the strands to rotate about each other in space. Some possibilities for the unwinding reaction are illustrated in **Figure 15.20**.

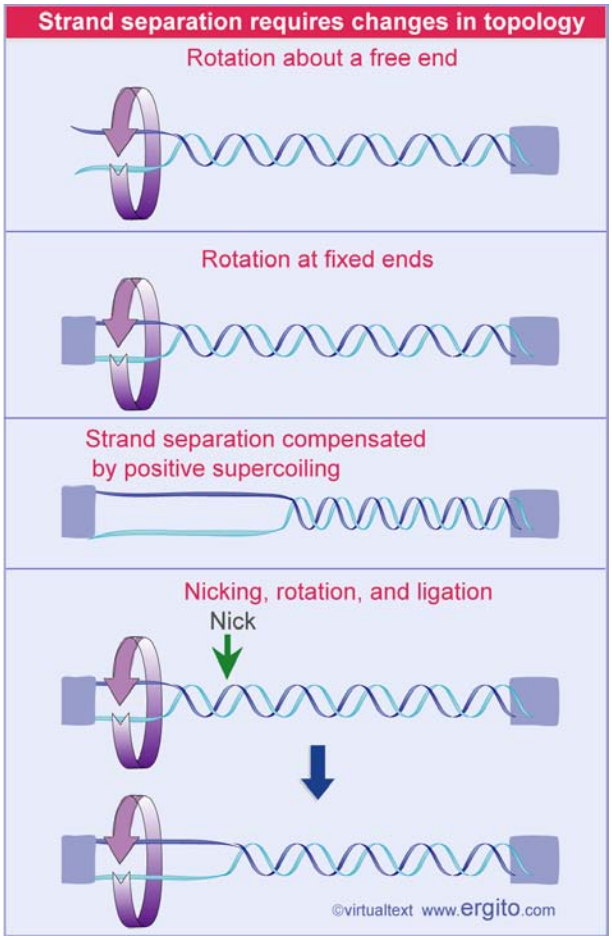

Figure 15.20 Separation of the strands of a DNA double helix could be achieved in several ways.

We might envisage the structure of DNA in terms of a free end that would allow the strands to rotate about the axis of the double helix for unwinding. Given the length of the double helix, however, this would involve the separating strands in a considerable amount of flailing about, which seems unlikely in the confines of the cell.

A similar result is achieved by placing an apparatus to control the rotation at the free end. However, the effect must be transmitted over a considerable distance, again involving the rotation of an unreasonable length of material.

Consider the effects of separating the two strands in a molecule whose ends are not free to rotate. When two intertwined strands are pulled apart from one end, the result is to increase their winding about each other farther along the molecule. The problem can be overcome by introducing a transient nick in one strand. An internal free end allows the nicked strand to rotate about the intact strand, after which the nick can be sealed. Each repetition of the nicking and sealing reaction releases one superhelical turn.

A closed molecule of DNA can be characterized by its **linking number**, the number

of times one strand crosses over the other in space. Closed DNA molecules of identical sequence may have different linking numbers, reflecting different degrees of supercoiling. Molecules of DNA that are the same except for their linking numbers are called **topological isomers**.

The linking number is made up of two components: the writhing number (W) and the twisting number (T).

The **twisting number**, T, is a property of the double helical structure itself, representing the rotation of one strand about the other. It represents the total number of turns of the duplex. It is determined by the number of base pairs per turn. For a relaxed closed circular DNA lying flat in a plane, the twist is the total number of base pairs divided by the number of base pairs per turn.

The **writhing number**, W, represents the turning of the axis of the duplex in space. It corresponds to the intuitive concept of supercoiling, but does not have exactly the same quantitative definition or measurement. For a relaxed molecule, W = 0, and the linking number equals the twist.

We are often concerned with the change in linking number, Δ L, given by the equation

$$\Delta L = \Delta W + \Delta T$$

The equation states that any change in the total number of revolutions of one DNA strand about the other can be expressed as the sum of the changes of the coiling of the duplex axis in space (Δ W) and changes in the screwing of the double helix itself (Δ T). In a free DNA molecule, W and T are freely adjustable, and any Δ L (change in linking number) is likely to be expressed by a change in W, that is, by a change in supercoiling.

A decrease in linking number, that is, a change of - Δ L, corresponds to the introduction of some combination of negative supercoiling and/or underwinding. An increase in linking number, measured as a change of + Δ L, corresponds to a decrease in negative supercoiling/underwinding.

We can describe the change in state of any DNA by the specific linking difference, $\sigma = \Delta L/L_0$, where L_0 is the linking number when the DNA is relaxed. If all of the change in linking number is due to change in W (that is, $\Delta T = 0$), the specific linking difference equals the supercoiling density. In effect, σ as defined in terms of $\Delta L/L_0$ can be assumed to correspond to superhelix density so long as the structure of the double helix itself remains constant.

The critical feature about the use of the linking number is that this parameter is an invariant property of any individual closed DNA molecule. The linking number cannot be changed by any deformation short of one that involves the breaking and rejoining of strands. A circular molecule with a particular linking number can express it in terms of different combinations of T and W, but cannot change their sum so long as the strands are unbroken. (In fact, the partition of L between T and W prevents the assignment of fixed values for the latter parameters for a DNA molecule in solution.)

The linking number is related to the actual enzymatic events by which changes are made in the topology of DNA. The linking number of a particular closed molecule can be changed only by breaking a strand or strands, using the free end to rotate one strand about the other, and rejoining the broken ends. When an enzyme performs such an action, it must change the linking number by an integer; this value can be determined as a characteristic of the reaction. Then we can consider the effects of this change in terms of ΔW and ΔT.

RECOMBINATION AND REPAIR

4.15.13 Topoisomerases relax or introduce supercoils in DNA

Key Terms

A **type I topoisomerase** is an enzyme that changes the topology of DNA by nicking and resealing one strand of DNA.

A **type II topoisomerase** is an enzyme that changes the topology of DNA by nicking and resealing both strands of DNA.

Key Concepts

- Topoisomerases change the linking number by breaking bonds in DNA, changing the conformation of the double helix in space, and remaking the bonds.
- Type I enzymes act by breaking a single strand of DNA; type II enzymes act by making double-strand breaks.

Changes in the topology of DNA can be caused in several ways.**Figure 15.21** shows some examples. In order to start replication or transcription, the two strands of DNA must be unwound. In the case of replication, the two strands separate permanently. and each reforms a duplex with the newly-synthesized daughter strand. In the case of transcription, the movement of RNA polymerase creates a region of positive supercoiling in front and a region of negative supercoiling behind the enzyme. This must be resolved before the positive supercoils impede the movement of the enzyme (see *Molecular Biology 3.9.15 Supercoiling is an important feature of transcription*). When a circular DNA molecule is replicated, the circular products may be catenated, with one passed through the other. They must be separated in order for the daughter molecules to segregate to separate daughter cells. Yet another situation in which supercoiling is important is the folding of the DNA thread into a chain of nucleosomes in the eukaryotic nucleus (see *Molecular Biology 5.20.6 The periodicity of DNA changes on the nucleosome*). All of the situations are resolved by the actions of topoisomerases.

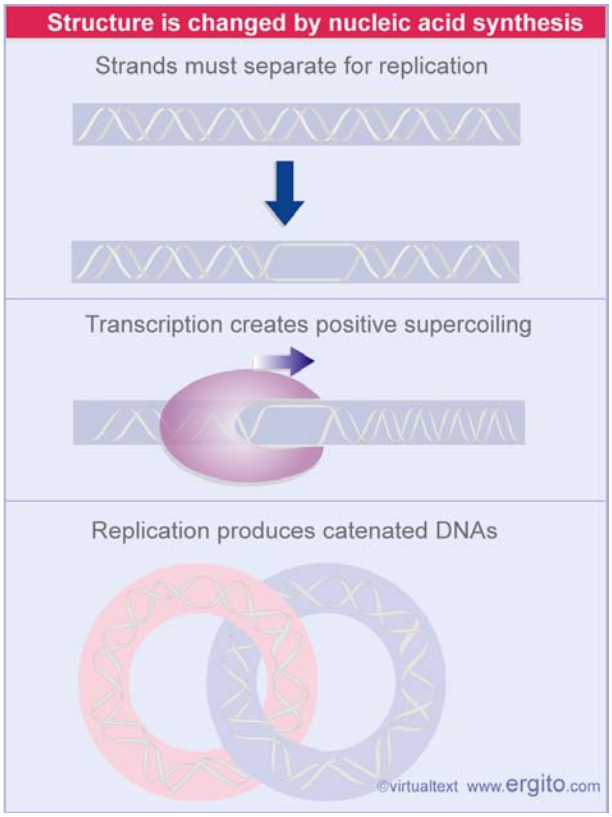

Figure 15.21 The topological structure of DNA is changed during replication and transcription. Strand separation for replication (or transcription) requires a base turn of DNA to be unwound. Transcription creates positive supercoils ahead of the RNA polymerase. Replication of a circular template produces two catenated daughter templates.

DNA topoisomerases are enzymes that catalyze changes in the topology of DNA by transiently breaking one or both strands of DNA, passing the unbroken strand(s) through the gap, and then resealing the gap (for review see 2280). The ends that are generated by the break are never free, but are manipulated exclusively within the confines of the enzyme – in fact, they are covalently linked to the enzyme. Topoisomerases act on DNA irrespective of its sequence, but some enzymes involved in site-specific recombination function in the same way and also fit the definition of topoisomerases (see *Molecular Biology 4.15.18 Site-specific recombination resembles topoisomerase activity*).

Topoisomerases are divided into two classes, according to the nature of the mechanisms they employ (for review see 3434). **Type I topoisomerases** act by making a transient break in one strand of DNA. **Type II topoisomerases** act by introducing a transient double-strand break. Topoisomerases in general vary with regard to the types of topological change they introduce. Some topoisomerases can relax (remove) only negative supercoils from DNA; others can relax both negative and positive supercoils. Enzymes that can introduce negative supercoils are called gyrases; those that can introduce positive supercoils are called reverse gyrases.

There are four topoisomerase enzymes in *E. coli*, called topoisomerase I, III, IV and

DNA gyrase. DNA topoisomerase I and III are type I enzymes. Gyrase and DNA topoisomerase IV are type II enzymes. Each of the four enzymes is important in one or more of the situations described in **Figure 15.21**:

- The overall level of negative supercoiling in the bacterial nucleoid is the result of a balance between the introduction of supercoils by gyrase and their relaxation by topoisomerases I and IV. This is a crucial aspect of nucleoid structure (see *Molecular Biology 5.19.4 The bacterial genome is supercoiled*), and affects initiation of transcription at certain promoters (see *Molecular Biology 3.9.15 Supercoiling is an important feature of transcription*).

- The same enzymes are involved in resolving the problems created by transcription; gyrase converts the positive supercoils that are generated ahead of RNA polymerase into negative supercoils, and topoisomerases I and IV remove the negative supercoils that are left behind the enzyme. Similar, but more complicated, effects occur during replication, and the enzymes have similar roles in dealing with them.

- As replication proceeds, the daughter duplexes can become twisted around one another, in a stage known as precatenation. The precatenanes are removed by topoisomerase IV, which also decatenates any catenated genomes that are left at the end of replication. The functions of topoisomerase III partially overlap those of topoisomerase IV.

The enzymes in eukaryotes follow the same principles, although the detailed division of responsibilities may be different. They do not show sequence or structural similarity with the prokaryotic enzymes. Most eukaryotes contain a single topoisomerase I enzyme that is required both for replication fork movement and for relaxing supercoils generated by transcription. A topoisomerase II enzyme(s) is required to unlink chromosomes following replication. Other individual topoisomerases have been implicated in recombination and repair activities.

Last updated on 5-14-2002

Reviews

2280. Champoux, J. J. (2001). *DNA topoisomerases: structure, function, and mechanism.* Annu. Rev. Biochem. 70, 369-413.

3434. Wang, J. C. (2002). *Cellular roles of DNA topoisomerases: a molecular perspective.* Nat. Rev. Mol. Cell Biol. 3, 430-440.

RECOMBINATION AND REPAIR

4.15.14 Topoisomerases break and reseal strands

Key Terms

Single-strand passage is a reaction catalyzed by type I topoisomerase in which one section of single-stranded DNA is passed through another strand.

A **knot** in the DNA is an entangled region that cannot be resolved without cutting and rearranging the DNA.

To **catenate** is to link together two circular molecules as in a chain.

Key Concepts

- Type I topoisomerases function by forming a covalent bond to one of the broken ends, moving one strand around the other, and then transferring the bound end to the other broken end. Because bonds are conserved, no input of energy is required.

The common action for all topoisomerases is to link one end of each broken strand to a tyrosine residue in the enzyme (for review see 2280). A type I enzyme links to the single broken strand; a type II enzyme links to one end of each broken strand. The topoisomerases are further divided into the A and B groups according to whether the linkage is to a 5′ phosphate or 3′ phosphate. The use of the transient phosphodiester-tyrosine bond suggests a mechanism for the action of the enzyme; it transfers a phosphodiester bond(s) in DNA to the protein, manipulates the structure of one or both DNA strands, and then rejoins the bond(s) in the original strand.

The *E. coli* enzymes are all of type A, using links to 5′ phosphate. This is the general pattern for bacteria, where there are almost no type B topoisomerases. All four possible types of topoisomerase (IA, IB, IIA, IIB) are found in eukaryotes.

A model for the action of topoisomerase IA is illustrated in **Figure 15.22**. The enzyme binds to a region in which duplex DNA becomes separated into its single strands; then it breaks one strand, pulls the other strand through the gap, and finally seals the gap. The transfer of bonds from nucleic acid to protein explains how the enzyme can function without requiring any input of energy. There has been no irreversible hydrolysis of bonds; their energy has been conserved through the transfer reactions. The model is supported by the crystal structure of the enzyme (2460).

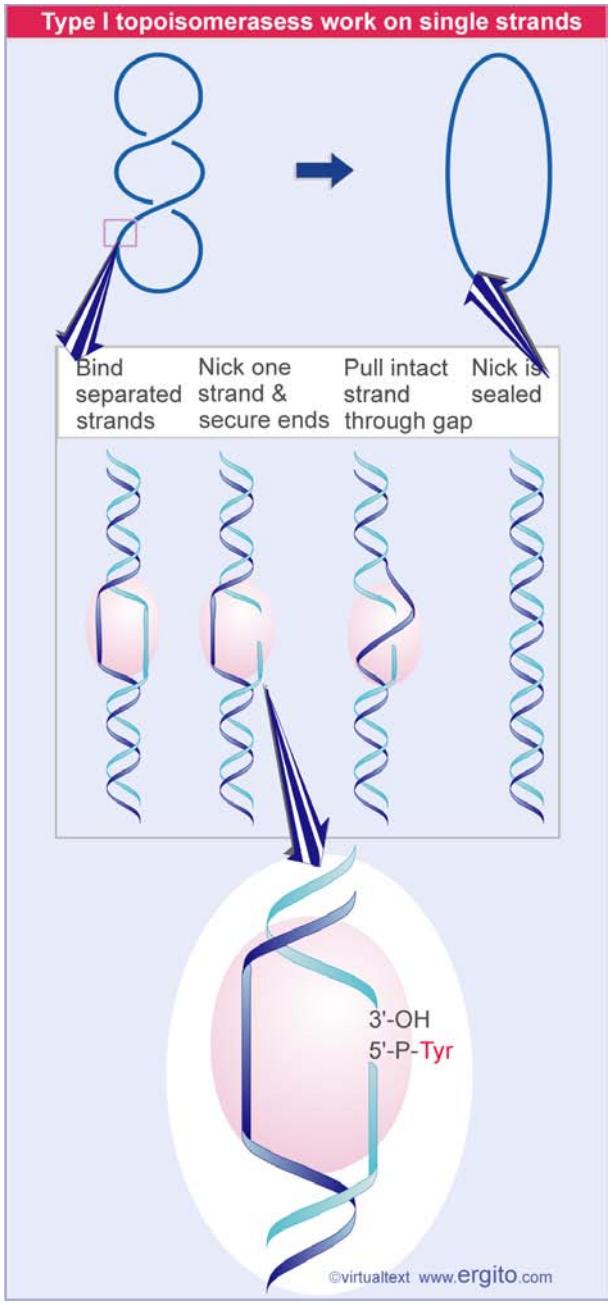

Figure 15.22 Bacterial type I topoisomerases recognize partially unwound segments of DNA and pass one strand through a break made in the other.

The reaction changes the linking number in steps of 1. Each time one strand is passed through the break in the other, there is a ΔL of +1. The figure illustrates the enzyme activity in terms of moving the individual strands. In a free supercoiled molecule, the interchangeability of W and T should let the change in linking number be taken up by a change of ΔW = +1, that is, by one less turn of negative supercoiling.

The reaction is equivalent to the rotation illustrated in the bottom part of **Figure**

15.20, with the restriction that the enzyme limits the reaction to a single-strand passage per event. (By contrast, the introduction of a nick in a supercoiled molecule allows free strand rotation to relieve all the tension by multiple rotations.)

The type I topoisomerase also can pass one segment of a single-stranded DNA through another. This **single-strand passage** reaction can introduce **knots** in DNA and can **catenate** two circular molecules so that they are connected like links on a chain. We do not understand the uses (if any) to which these reactions are put *in vivo*.

Type II topoisomerases generally relax both negative and positive supercoils. The reaction requires ATP, with one ATP hydrolyzed for each catalytic event. As illustrated in **Figure 15.23**, the reaction is mediated by making a double-stranded break in one DNA duplex. The double-strand is cleaved with a 4-base stagger between the ends, and each subunit of the dimeric enzyme attaches to a protruding broken end. Then another duplex region is passed through the break. The ATP is used in the following religation/release step, when the ends are rejoined and the DNA duplexes are released. This is why inhibiting the ATPase activity of the enzyme results in a "cleavable complex" that contains broken DNA.

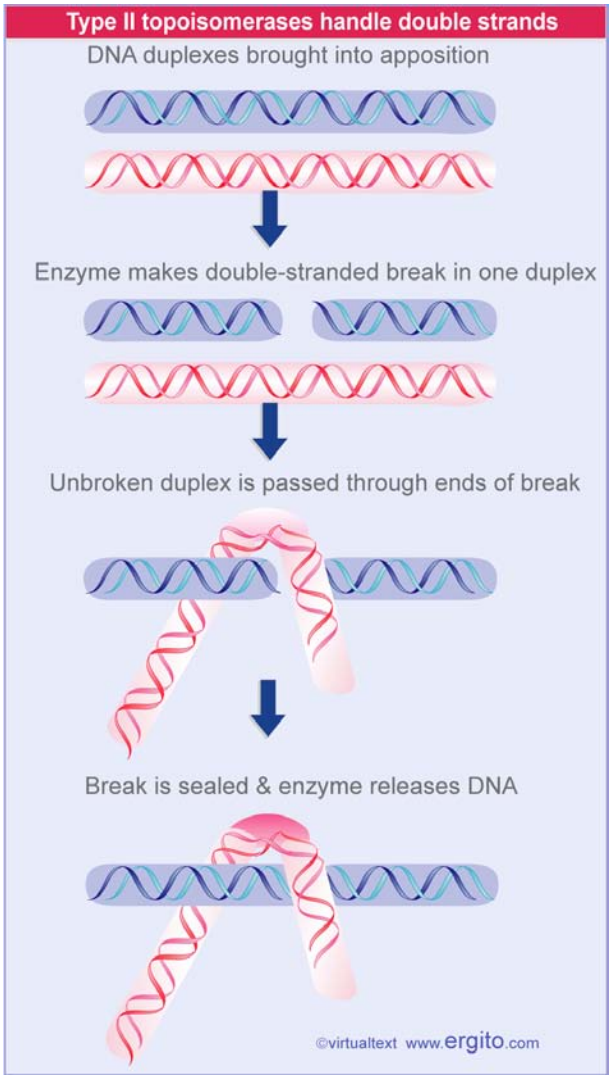

Figure 15.23 Type II topoisomerases can pass a duplex DNA through a double-strand break in another duplex.

A formal consequence of two-strand transfer is that the linking number is always changed in multiples of two. The topoisomerase II activity can be used also to introduce or resolve catenated duplex circles and knotted molecules.

The reaction probably represents a nonspecific recognition of duplex DNA in which the enzyme binds any two double-stranded segments that cross each other. The hydrolysis of ATP may be used to drive the enzyme through conformational changes that provide the force needed to push one DNA duplex through the break made in the other. Because of the topology of supercoiled DNA, the relationship of the crossing segments allows supercoils to be removed from either positively or negatively supercoiled circles.

Last updated on 5-14-2002

Reviews

2280. Champoux, J. J. (2001). *DNA topoisomerases: structure, function, and mechanism.* Annu. Rev. Biochem. 70, 369-413.

References

2460. Lima, C. D., Wang, J. C., and Mondragon, A. (1994). *Three-dimensional structure of the 67K N-terminal fragment of E. coli DNA topoisomerase I.* Nature 367, 138-146.

RECOMBINATION AND REPAIR

4.15.15 Gyrase functions by coil inversion

Key Terms

The **sign inversion** model describes the mechanism of DNA gyrase. DNA gyrase binds a positive supercoil (inducing a compensatory negative supercoil elsewhere on the closed circular DNA), breaks both strands in one duplex, passes the other duplex through, and reseals the strands.

Enzyme turnover is the process through which the enzyme returns to its original shape, enabling the enzyme to catalyze another reaction.

Key Concepts

- *E. coli* gyrase is a type II topoisomerase that used hydrolysis of ATP to provide energy to introduce negative supercoils into DNA.

Bacterial DNA gyrase is a topoisomerase of type II that is able to introduce negative supercoils into a relaxed closed circular molecule. DNA gyrase binds to a circular DNA duplex and supercoils it processively and catalytically: it continues to introduce supercoils into the same DNA molecule. One molecule of DNA gyrase can introduce ~100 supercoils per minute.

The supercoiled form of DNA has a higher free energy than the relaxed form, and the energy needed to accomplish the conversion is supplied by the hydrolysis of ATP. In the absence of ATP, the gyrase can relax negative but not positive supercoils, although the rate is more than 10× slower than the rate of introducing supercoils.

The *E. coli* DNA gyrase is a tetramer consisting of two types of subunit, each of which is a target for antibiotics (the most often used being nalidixic acid which acts on GyrA, and novobiocin which acts on GyrB). The drugs inhibit replication, which suggests that DNA gyrase is necessary for DNA synthesis to proceed. Mutations that confer resistance to the antibiotics identify the loci that code for the subunits.

Gyrase binds its DNA substrate around the outside of the protein tetramer. Gyrase protects ~140 bp of DNA from digestion by micrococcal nuclease. The **sign inversion** model for gyrase action is illustrated in **Figure 15.24**. The enzyme binds the DNA in a crossover configuration that is equivalent to a positive supercoil. This induces a compensating negative supercoil in the unbound DNA. Then the enzyme breaks the double strand at the crossover of the positive supercoil, passes the other duplex through, and seals the break.

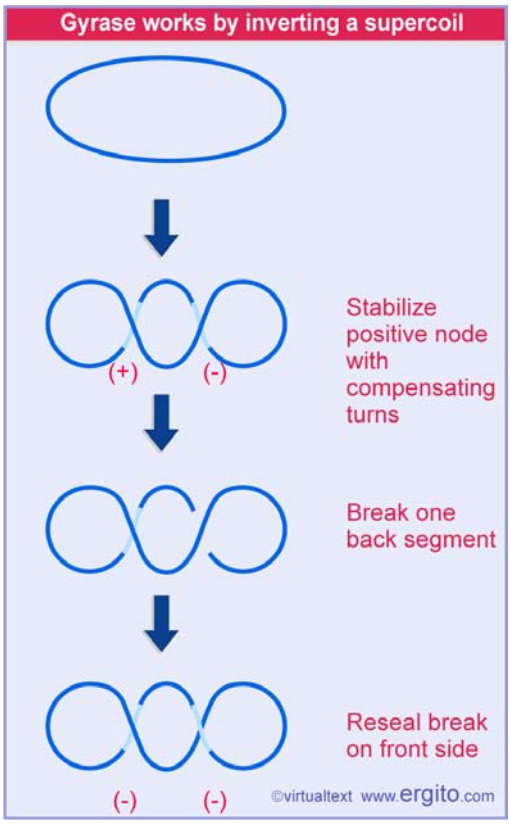

Figure 15.24 DNA gyrase may introduce negative supercoils in duplex DNA by inverting a positive supercoil.

The reaction directly inverts the sign of the supercoil: it has been converted from a +1 turn to a –1 turn. So the linking number has changed by $\Delta L = -2$, conforming with the demand that all events involving double-strand passage must change the linking number by a multiple of two.

Gyrase then releases one of the crossing segments of the (now negative) bound supercoil; this allows the negative turns to redistribute along DNA (as change in either T or W or both), and the cycle begins again. The same type of topological manipulation is responsible for catenation and knotting.

On releasing the inverted supercoil, the conformation of gyrase changes. For the enzyme to undertake another cycle of supercoiling, its original conformation must be restored. This process is called **enzyme turnover**. It is thought to be driven by the hydrolysis of ATP, since the replacement of ATP by an analog that cannot be hydrolyzed allows gyrase to introduce only one inversion (–2 supercoils) per substrate. So the enzyme does not need ATP for the supercoiling reaction, but does need it to undertake a second cycle. Novobiocin interferes with the ATP-dependent reactions of gyrase, by preventing ATP from binding to the B subunit.

The ATP-independent relaxation reaction is inhibited by nalidixic acid. This implicates the A subunit in the breakage and reunion reaction. Treating gyrase with nalidixic acid allows DNA to be recovered in the form of fragments generated by a

staggered cleavage across the duplex. The termini all possess a free 3′–OH group and a 4-base 5′ single-strand extension covalently linked to the A subunit. The covalent linkage retains the energy of the phosphate bond; this can be used to drive the sealing reaction, explaining why gyrase can undertake relaxation without ATP. The sites of cleavage are fairly specific, occurring about once every 100 bp.

Last updated on 5-14-2002

RECOMBINATION AND REPAIR

4.15.16 Specialized recombination involves specific sites

Key Terms

Site-specific recombination (Specialized recombination) occurs between two specific sequences, as in phage integration/excision or resolution of cointegrate structures during transposition.

Prophage is a phage genome covalently integrated as a linear part of the bacterial chromosome.

Integration of viral or another DNA sequence describes its insertion into a host genome as a region covalently linked on either side to the host sequences.

The **excision** step in an excision-repair system consists of removing a single-stranded stretch of DNA by the action of a 5′-3′ exonuclease.

A **secondary attachment site** is a locus on the bacterial chromosome into which phage lambda integrate inefficiently because the site resembles the *att* site.

The **core** sequence is the segment of DNA that is common to the attachment sites on both the phage lambda and bacterial genomes. It is the location of the recombination event that allows phage lambda to integrate.

The **arms** of a lambda phage attachment site are the sequences flanking the core region where the recombination event occurs.

Key Concepts

- Specialized recombination involves reaction between specific sites that are not necessarily homologous.
- Phage lambda integrates into the bacterial chromosome by recombination between a site on the phage and the *att* site on the *E. coli* chromosome.
- The phage is excised from the chromosome by recombination between the sites at the end of the linear prophage.
- Phage lambda *int* codes for an integrase that catalyzes the integration reaction.

Specialized recombination involves a reaction between two specific sites. The target sites are short, typically in a 14-50 bp length range. In some cases the two sites have the same sequence, but in other cases they are nonhomologous. The reaction is used to insert a free phage DNA into the bacterial chromosome or to excise an integrated phage DNA from the chromosome, and in this case the two recombining sequences are different from one another. It is also used before division to regenerate monomeric circular chromosomes from a dimer that has been created by a generalized recombination event (see *Molecular Biology 4.13.19 Chromosomal segregation may require site-specific recombination*). In this case the recombining sequences are identical.

The enzymes that catalyze site-specific recombination are generally called recombinases, and >100 of them are now known (2921). Those involved in phage integration or related to these enzymes are also known as the integrase family. Prominent members of the integrase family are the prototype Int from phage lambda, Cre from phage P1, and the yeast FLP enzyme (which catalyzes a chromosomal inversion) (for review see 2922).

The classic model for **site-specific recombination** is illustrated by phage lambda. The conversion of lambda DNA between its different life forms involves two types of event. The pattern of gene expression is regulated as described in *Molecular Biology 3.12 Phage strategies*. And the physical condition of the DNA is different in the lysogenic and lytic states:

- In the lytic lifestyle, lambda DNA exists as an independent, circular molecule in the infected bacterium.
- In the lysogenic state, the phage DNA is an integral part of the bacterial chromosome (called **prophage**).

Transition between these states involves site-specific recombination:

- To enter the lysogenic condition, free lambda DNA must be inserted into the host DNA. This is called **integration**.
- To be released from lysogeny into the lytic cycle, prophage DNA must be released from the chromosome. This is called **excision**.

Integration and excision occur by recombination at specific loci on the bacterial and phage DNAs called attachment (**att**) sites. The attachment site on the bacterial chromosome is called *att* λ in bacterial genetics. The locus is defined by mutations that prevent integration of lambda; it is occupied by prophage λ in lysogenic strains. When the *att* λ site is deleted from the *E. coli* chromosome, an infecting lambda phage can establish lysogeny by integrating elsewhere, although the efficiency of the reaction is <0.1% of the frequency of integration at *att* λ. This inefficient integration occurs at **secondary attachment sites**, which resemble the authentic *att* sequences.

For describing the integration/excision reactions, the bacterial attachment site (*att* λ) is called *attB*, consisting of the sequence components *BOB′*. The attachment site on the phage, *attP*, consists of the components *POP′*. **Figure 15.25** outlines the recombination reaction between these sites. The sequence O is common to *attB* and *attP*. It is called the **core** sequence; and the recombination event occurs within it. The flanking regions *B*, *B′* and *P*, *P′* are referred to as the **arms**; each is distinct in sequence. Because the phage DNA is circular, the recombination event inserts it into the bacterial chromosome as a linear sequence. The prophage is bounded by two new *att* sites, the products of the recombination, called *attL* and *attR*.

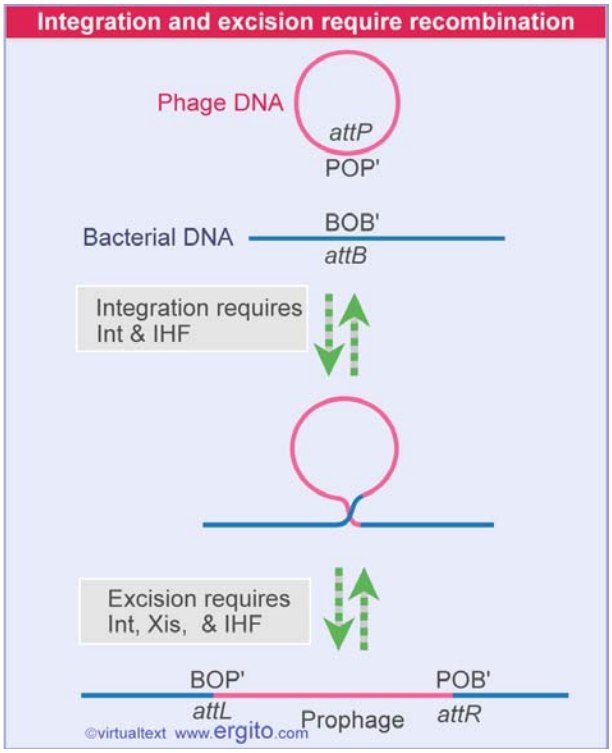

Figure 15.25 Circular phage DNA is converted to an integrated prophage by a reciprocal recombination between *attP* and *attB*; the prophage is excised by reciprocal recombination between *attL* and *attR*.

An important consequence of the constitution of the *att* sites is that the integration and excision reactions do not involve the same pair of reacting sequences. Integration requires recognition between *attP* and *attB*; while excision requires recognition between *attL* and *attR*. The directional character of site-specific recombination is controlled by the identity of the recombining sites.

Although the recombination event is reversible, different conditions prevail for each direction of the reaction. This is an important feature in the life of the phage, since it offers a means to ensure that an integration event is not immediately reversed by an excision, and vice versa.

The difference in the pairs of sites reacting at integration and excision is reflected by a difference in the proteins that mediate the two reactions:

- Integration (*attB*×*attP*) requires the product of the phage gene *int*, which codes for an integrase enzyme, and a bacterial protein called integration host factor (IHF).
- Excision (*attL*×*attR*) requires the product of phage gene *xis*, in addition to Int and IHF.

So Int and IHF are required for both reactions. Xis plays an important role in

controlling the direction; it is required for excision, but inhibits integration.

A similar system, but with somewhat simpler requirements for both sequence and protein components, is found in the bacteriophage P1. The Cre recombinase coded by the phage catalyzes a recombination between two target sequences. Unlike phage lambda, where the recombining sequences are different, in phage P1 they are identical. Each consists of a 34 bp-long sequence called *loxP*. The Cre recombinase is sufficient for the reaction; no accessory proteins are required. Because of its simplicity and its efficiency, what is now known as the Cre/*lox* system has been adapted for use in eukaryotic cells, where it has become one of the standard techniques for undertaking site-specific recombination (2923).

Last updated on September 11, 2002

Reviews

2922. Craig, N. L. (1988). *The mechanism of conservative site-specific recombination.* Annu. Rev. Genet. 22, 77-105.

References

2921. Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T., and Landy, A. (1998). *Similarities and differences among 105 members of the Int family of site-specific recombinases.* Nucleic Acids Res. 26, 391-406.

2923. Metzger, D., Clifford, J., Chiba, H., and Chambon, P. (1995). *Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase.* Proc. Natl. Acad. Sci. USA 92, 6991-6995.

RECOMBINATION AND REPAIR

4.15.17 Site-specific recombination involves breakage and reunion

Key Concepts

- Cleavages staggered by 7 bp are made in both *attB* and *attP* and the ends are joined cross-wise.

The *att* sites have distinct sequence requirements, and *attP* is much larger than *attB*. The function of *attP* requires a stretch of 240 bp, but the function of *attB* can be exercised by the 23 bp fragment extending from –11 to +11, in which there are only 4 bp on either side of the core. The disparity in their sizes suggests that *attP* and *attB* play different roles in the recombination, with *attP* providing additional information necessary to distinguish it from *attB*.

Does the reaction proceed by a concerted mechanism in which the strands in *attP* and *attB* are cut simultaneously and exchanged? Or are the strands exchanged one pair at a time, the first exchange generating a Holliday junction, the second cycle of nicking and ligation occurring to release the structure? The alternatives are depicted in **Figure 15.26**.

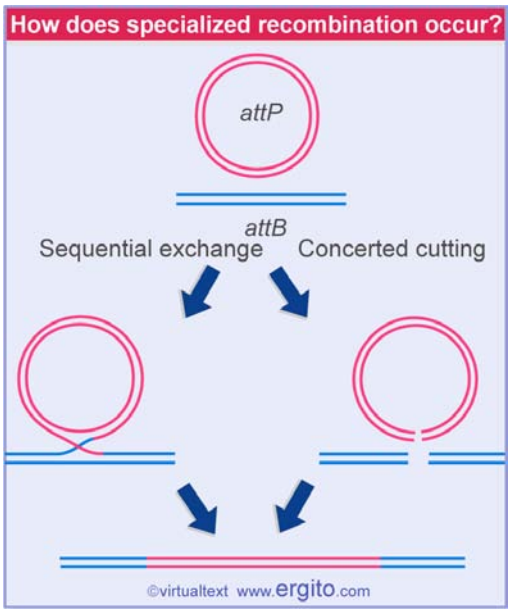

Figure 15.26 Does recombination between *attP* and *attB* proceed by sequential exchange or concerted cutting?

The recombination reaction has been halted at intermediate stages by the use of "suicide substrates," in which the core sequence is nicked. The presence of the nick interferes with the recombination process. This makes it possible to identify

molecules in which recombination has commenced but has not been completed. The structures of these intermediates suggest that exchanges of single strands take place sequentially.

The model illustrated in **Figure 15.27** shows that if *attP* and *attB* sites each suffer the same staggered cleavage, complementary single-stranded ends could be available for crosswise hybridization. The distance between the lambda crossover points is 7 bp, and the reaction generates 3′–phosphate and 5′–OH ends. The reaction is shown for simplicity as generating overlapping single-stranded ends that anneal, but actually occurs by a process akin to the recombination event of **Figure 15.5**. The corresponding strands on each duplex are cut at the same position, the free 3′ ends exchange between duplexes, the branch migrates for a distance of 7 bp along the region of homology, and then the structure is resolved by cutting the other pair of corresponding strands.

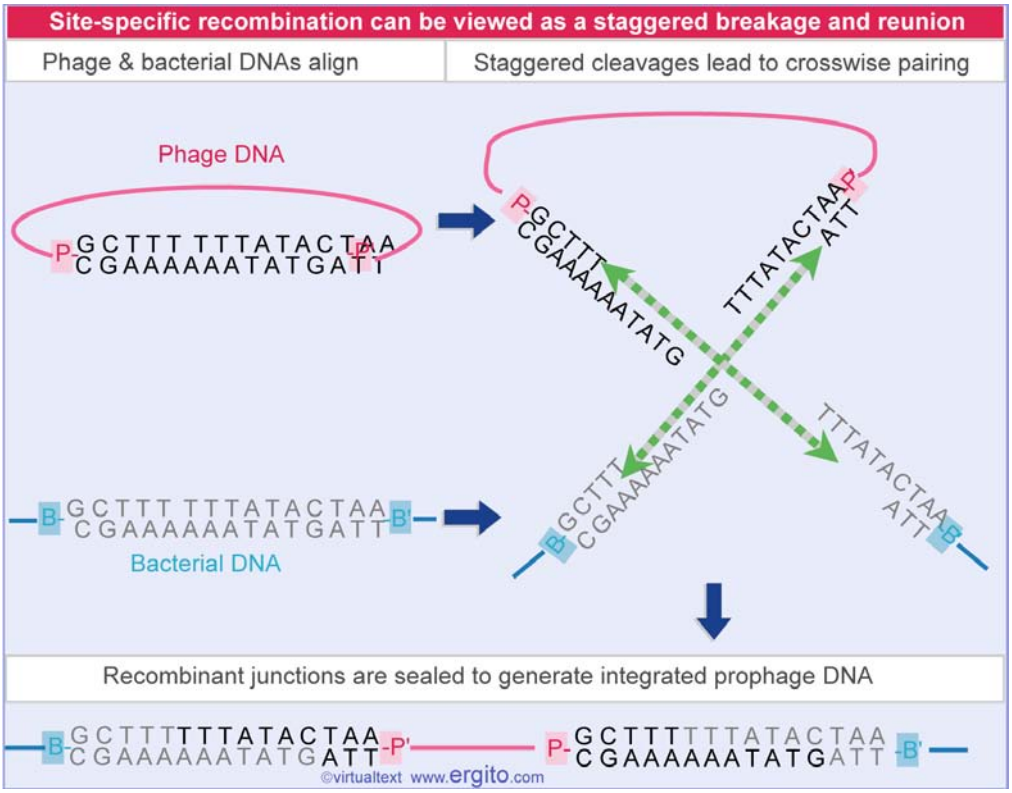

Figure 15.27 Staggered cleavages in the common core sequence of *attP* and *attB* allow crosswise reunion to generate reciprocal recombinant junctions.

Last updated on 9-11-2002

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.4.15.17

RECOMBINATION AND REPAIR

4.15.18 Site-specific recombination resembles topoisomerase activity

Key Concepts

- Integrases are related to topoisomerases, and the recombination reaction resembles topoisomerase action except that nicked strands from *different* duplexes are sealed together.
- The reaction conserves energy by using a catalytic tyrosine in the enzyme to break a phosphodiester bond and link to the broken 3′ end.
- Two enzyme units bind to each recombination site and the two dimers synapse to form a complex in which the transfer reactions occur.

Integrases use a mechanism similar to that of type I topoisomerases, in which a break is made in one DNA strand at a time. The difference is that a recombinase reconnects the ends cross-wise, whereas a topoisomerase makes a break, manipulates the ends, and then rejoins the original ends. The basic principle of the system is that four molecules of the recombinase are required, one to cut each of the four strands of the two duplexes that are recombining.

Figure 15.28 shows the nature of the reaction catalyzed by an integrase. The enzyme is a monomeric protein that has an active site capable of cutting and ligating DNA. The reaction involves an attack by a tyrosine on a phosphodiester bond. The 3′ end of the DNA chain is linked through a phosphodiester bond to a tyrosine in the enzyme. This releases a free 5′ hydroxyl end.

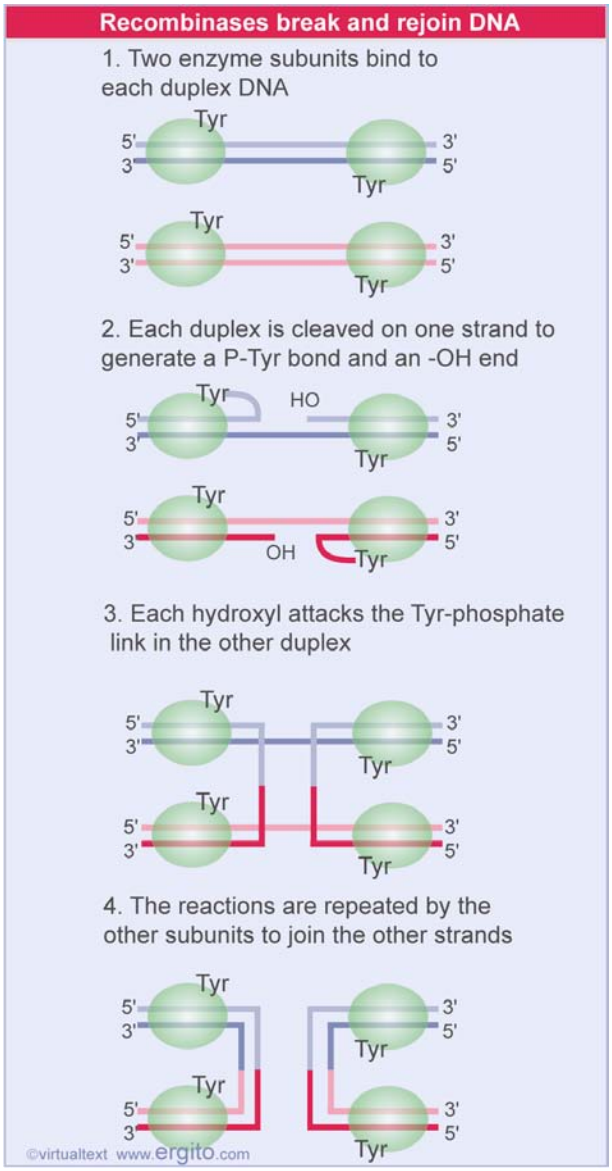

Figure 15.28 Integrases catalyze recombination by a mechanism similar to topoisomerases. Staggered cuts are made in DNA and the 3′-phosphate end is covalently linked to a tyrosine in the enzyme. Then the free hydroxyl group of each strand attacks the P-Tyr link of the other strand. The first exchange shown in the figure generates a Holliday structure. The structure is resolved by repeating the process with the other pair of strands.

Two enzyme units are bound to each of the recombination sites. At each site, only one of the enzyme active sites attacks the DNA. The symmetry of the system ensures that complementary strands are broken in each recombination site. Then the free 5′–OH end in each site attacks the 3′–phosphotyrosine link in the other site. This generates a Holliday junction.

The structure is resolved when the other two enzyme units (which had not been

involved in the first cycle of breakage and reunion) act on the other pair of complementary strands.

The successive interactions accomplish a conservative strand exchange, in which there are no deletions or additions of nucleotides at the exchange site, and there is no need for input of energy. The transient 3′–phospho-tyrosine link between protein and DNA conserves the energy of the cleaved phosphodiester bond.

Figure 15.29 shows the reaction intermediate, based on the crystal structure (2924). (Trapping the intermediate was made possible by using a "suicide substrate", consisting of a synthetic DNA duplex with a missing phosphodiester bond, so that the attack by the enzyme does not generate a free 5′–OH end.) The structure of the Cre-*lox* complex shows two Cre molecules, each bound to a 15 bp length of DNA. The DNA is bent by ~100° at the center of symmetry. Two of these complexes assemble in an anti-parallel way to form a tetrameric protein structure bound to two synapsed DNA molecules. Strand exchange takes place in a central cavity of the protein structure that contains the central 6 bases of the crossover region.

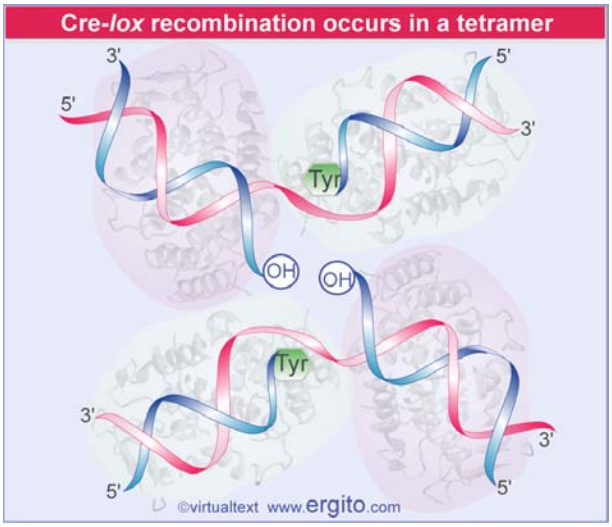

Figure 15.29 A synapsed *loxA* recombination complex has a tetramer of Cre recombinases, with one enzyme monomer bound to each half site. Two of the four active sites are in use, acting on complementary strands of the two DNA sites.

The tyrosine that is responsible for cleaving DNA in any particular half site is provided by the enzyme subunit that is bound to that half site. This is called *cis* cleavage. This is true also for the Int integrase and XerD recombinase. However, the FLP recombinase cleaves in *trans*, involving a mechanism in which the enzyme subunit that provides the tyrosine is *not* the subunit bound to that half site, but is one of the other subunits.

Last updated on 9-11-2002

References

2924. Guo, F., Gopaul, D. N., and van Duyne, G. D. (1997). *Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse.* Nature 389, 40-46.

RECOMBINATION AND REPAIR

4.15.19 Lambda recombination occurs in an intasome

Key Terms

An **intasome** is a protein-DNA complex between the phage lambda integrase (Int) and the phage lambda attachment site (*attP*).

Key Concepts

- Lambda integration takes place in a large complex that also includes the host protein IHF.
- The excision reaction requires Int and Xis and recognizes the ends of the prophage DNA as substrates.

Unlike the Cre/*lox* recombination system, which requires only the enzyme and the two recombining sites, phage lambda recombination occurs in a large structure, and has different components for each direction of the reaction (integration versus excision).

A host protein called IHF is required for both integration and excision. IHF is a 20 kD protein of two different subunits, coded by the genes *himA* and *himD*. IHF is not an essential protein in *E. coli*, and is not required for homologous bacterial recombination. It is one of several proteins with the ability to wrap DNA on a surface. Mutations in the *him* genes prevent lambda site-specific recombination, and can be suppressed by mutations in λ*int*, which suggests that IHF and Int interact. Site-specific recombination can be performed *in vitro* by Int and IHF.

The *in vitro* reaction requires supercoiling in *attP*, but not in *attB*. When the reaction is performed *in vitro* between two supercoiled DNA molecules, almost all of the supercoiling is retained by the products. So there cannot be any free intermediates in which strand rotation could occur. This was one of the early hints that the reaction proceeds through a Holliday junction. We now know that the reaction proceeds by the mechanism typical of this class of enzymes, related to the topoisomerase I mechanism (see *Molecular Biology 4.15.18 Site-specific recombination resembles topoisomerase activity*).

Int has two different modes of binding. The C-terminal domain behaves like the Cre recombinase. It binds to inverted sites at the core sequence, positioning itself to make the cleavage and ligation reactions on each strand at the positions illustrated in **Figure 15.30**. The N-terminal domain binds to sites in the arms of *attP* that have a different consensus sequence. This binding is responsible for the aggregation of subunits into the intasome (2513). The two domains probably bind DNA simultaneously, thus bringing the arms of *attP* close to the core.

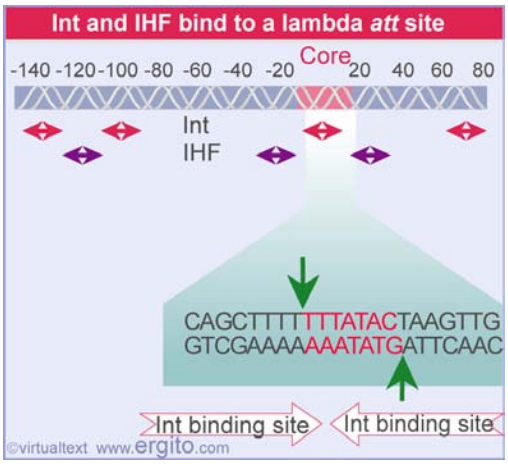

Figure 15.30 Int and IHF bind to different sites in *attP*. The Int recognition sequences in the core region include the sites of cutting.

IHF binds to sequences of ~20 bp in *attP*. The IHF binding sites are approximately adjacent to sites where Int binds. Xis binds to two sites located close to one another in *attP*, so that the protected region extends over 30-40 bp. Together, Int, Xis, and IHF cover virtually all of *attP*. The binding of Xis changes the organization of the DNA so that it becomes inert as a substrate for the integration reaction.

When Int and IHF bind to *attP*, they generate a complex in which all the binding sites are pulled together on the surface of a protein. Supercoiling of *attP* is needed for the formation of this **intasome**. The only binding sites in *attB* are the two Int sites in the core. But Int does not bind directly to *attB* in the form of free DNA. The intasome is the intermediate that "captures" *attB*, as indicated schematically in **Figure 15.31**.

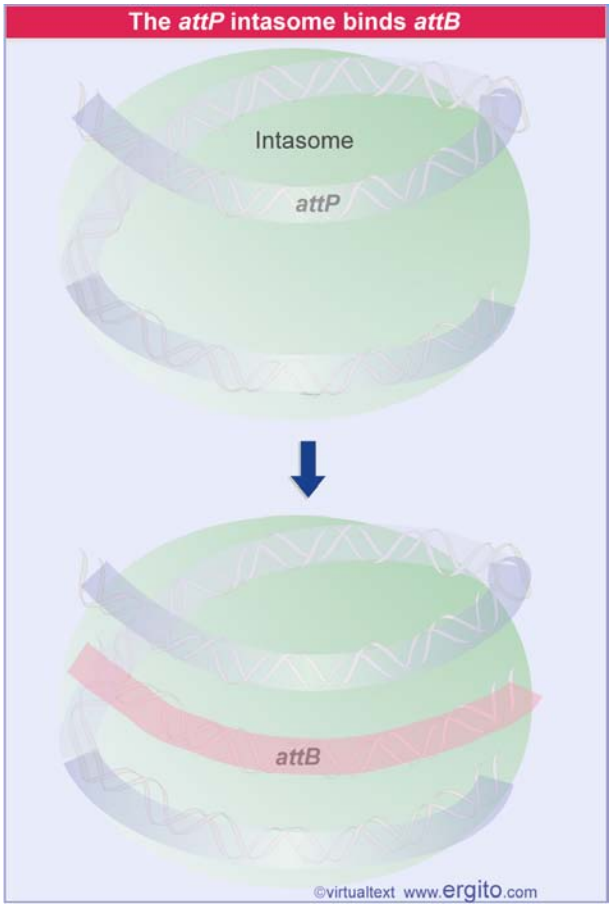

Figure 15.31 Multiple copies of Int protein may organize *attP* into an intasome, which initiates site-specific recombination by recognizing *attB* on free DNA.

According to this model, the initial recognition between *attP* and *attB* does not depend directly on DNA homology, but instead is determined by the ability of Int proteins to recognize both *att* sequences. The two *att* sites then are brought together in an orientation predetermined by the structure of the intasome. Sequence homology becomes important at this stage, when it is required for the strand exchange reaction.

The asymmetry of the integration and excision reactions is shown by the fact that Int can form a similar complex with *attR* only if Xis is added. This complex can pair with a condensed complex that Int forms at *attL*. IHF is not needed for this reaction.

Much of the complexity of site-specific recombination may be caused by the need to regulate the reaction so that integration occurs preferentially when the virus is entering the lysogenic state, while excision is preferred when the prophage is entering the lytic cycle. By controlling the amounts of Int and Xis, the appropriate reaction will occur.

Last updated on 9-11-2002

References

2513. Wojciak, J. M., Sarkar, D., Landy, A., and Clubb, R. T. (2002). *Arm-site binding by lambda -integrase: solution structure and functional characterization of its amino-terminal domain.* Proc. Natl. Acad. Sci. USA 99, 3434-3439.

RECOMBINATION AND REPAIR

4.15.20 Repair systems correct damage to DNA

Key Terms

A **structural distortion** is a change in the shape of a molecule.

A **pyrimidine dimer** is formed when ultraviolet irradiation generates a covalent link directly between two adjacent pyrimidine bases in DNA. It blocks DNA replication.

Photoreactivation uses a white-light-dependent enzyme to split cyclobutane pyrimidine dimers formed by ultraviolet light.

Mismatch repair corrects recently inserted bases that do not pair properly. The process preferentially corrects the sequence of the daughter strand by distinguishing the daughter strand and parental strand, sometimes on the basis of their states of methylation.

Excision repair describes a type of repair system in which one strand of DNA is directly excised and then replaced by resynthesis using the complementary strand as template.

Recombination-repair is a mode of filling a gap in one strand of duplex DNA by retrieving a homologous single strand from another duplex.

Error-prone synthesis occurs when DNA incorporates noncomplementary bases into the daughter strand.

Key Concepts

- Repair systems recognize DNA sequences that do not conform to standard base pairs.
- Excision systems remove one strand of DNA at the site of damage and then replace it.
- Recombination-repair systems use recombination to replace the double-stranded region that has been damaged.
- All these systems are prone to introducing errors during the repair process.
- Photoreactivation is a nonmutagenic repair system that acts specifically on pyrimidine dimers.

Any event that introduces a deviation from the usual double-helical structure of DNA is a threat to the genetic constitution of the cell. We can divide such changes into two general classes:

- *Single base changes* affect the sequence but not the overall structure of DNA. They do not affect transcription or replication, when the strands of the DNA duplex are separated. So these changes exert their damaging effects on future

generations through the consequences of the change in DNA sequence. The cause of this type of effect is the conversion of one base into another that is not properly paired with the partner base. They may be happen as the result of mutation of a base *in situ* or by replication errors. **Figure 15.32** shows that deamination of cytosine to uracil (spontaneously or by chemical mutagen) creates a mismatched U·G pair. **Figure 15.33** shows that a replication error might insert adenine instead of cytosine to create an A·G pair. Similar consequences could result from covalent addition of a small group to a base that modifies its ability to base pair. These changes may result in very minor structural distortion (as in the case of a U·G pair) or quite significant change (as in the case of an A·G pair), but the common feature is that the mismatch persists only until the next replication. So only limited time is available to repair the damage before it is fixed by replication.

- **Structural distortions** may provide a physical impediment to replication or transcription. Introduction of covalent links between bases on one strand of DNA or between bases on opposite strands inhibits replication and transcription. **Figure 15.34** shows the example of ultraviolet irradiation, which introduces covalent bonds between two adjacent thymine bases, giving an intrastrand **pyrimidine dimer**. **Figure 15.35** shows that similar consequences could result from addition of a bulky adduct to a base that distorts the structure of the double helix. A single-strand nick or the removal of a base, as shown in **Figure 15.36**, prevents a strand from serving as a proper template for synthesis of RNA or DNA. The common feature in all these changes is that the damaged adduct remains in the DNA, continuing to cause structural problems and/or induce mutations, until it is removed.

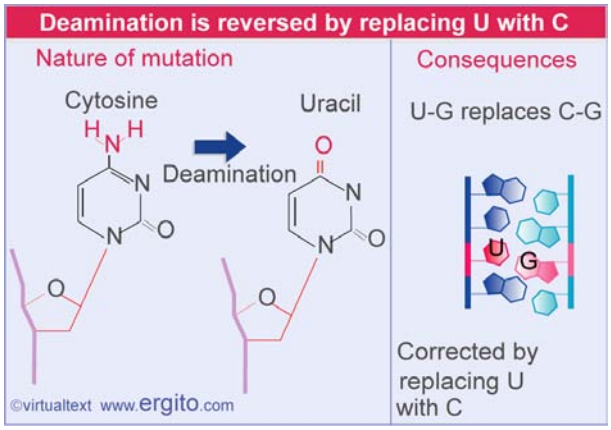

Figure 15.32 Deamination of cytosine creates a U-G base pair. Uracil is preferentially removed from the mismatched pair.

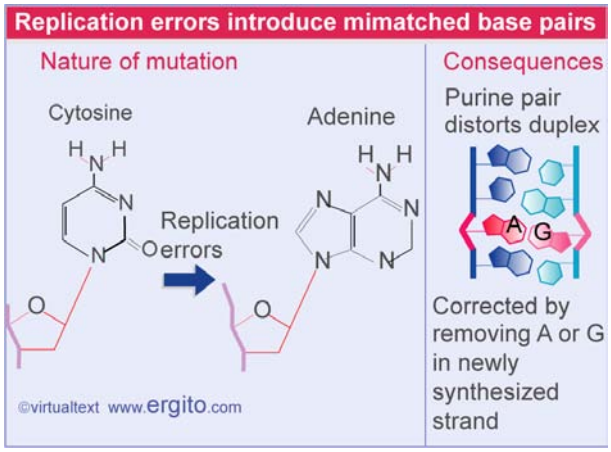

Figure 15.33 A replication error creates a mismatched pair that may be corrected by replacing one base; if uncorrected, a mutation is fixed in one daughter duplex.

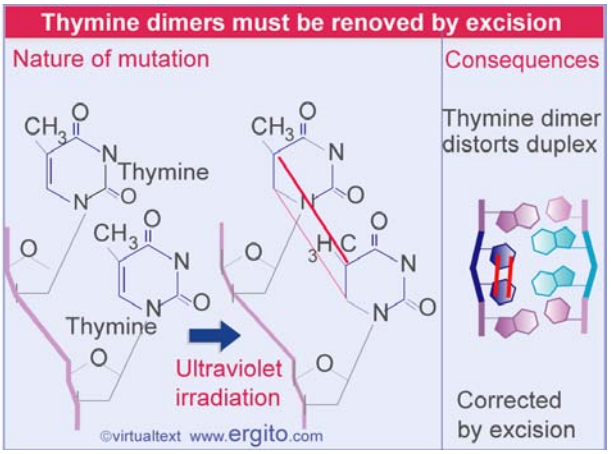

Figure 15.34 Ultraviolet irradiation causes dimer formation between adjacent thymines. The dimer blocks replication and transcription.

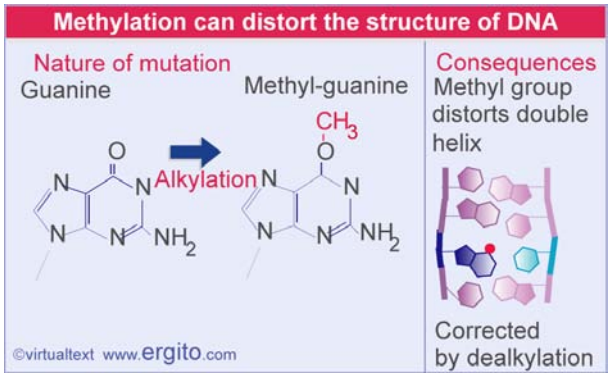

Figure 15.35 Methylation of a base distorts the double helix and causes mispairing at replication.

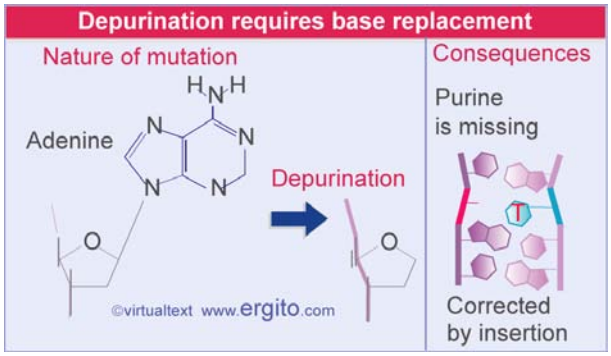

Figure 15.36 Depurination removes a base from DNA, blocking replication and transcription.

Injury to DNA is minimized by systems that recognize and correct the damage. The repair systems are as complex as the replication apparatus itself, which indicates their importance for the survival of the cell. When a repair system reverses a change to DNA, there is no consequence. But a mutation may result when it fails to do so. The measured rate of mutation reflects a balance between the number of damaging events occurring in DNA and the number that have been corrected (or miscorrected).

Repair systems often can recognize a range of distortions in DNA as signals for action, and a cell is likely to have several systems able to deal with DNA damage. The importance of DNA repair in eukaryotes is indicated by the identification of >130 repair genes in the human genome (for review see 2398). We may divide them into several general types, as summarized in **Figure 15.37**:

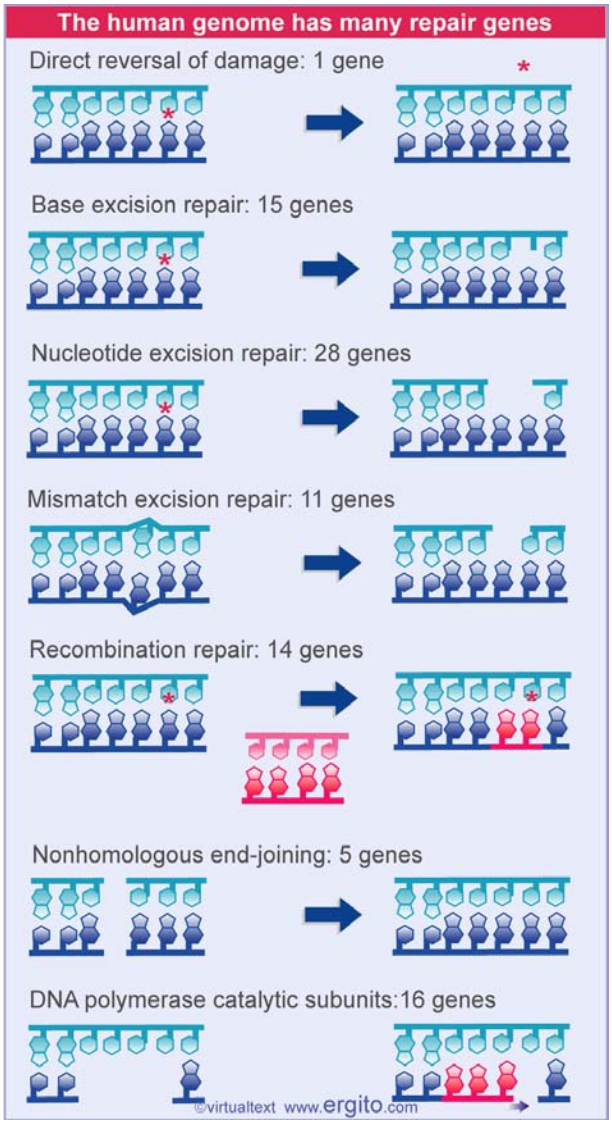

Figure 15.37 Repair genes can be classified into pathways that use different mechanisms to reverse or bypass damage to DNA.

- Some enzymes directly reverse specific sorts of damage to DNA.
- There are pathways for base excision repair, nucleotide excision repair, and mismatch repair, all of which function by removing and replacing material.
- There are systems that function by using recombination to retrieve an undamaged copy that is used to replace a damaged duplex sequence.
- The nonhomologous end-joining pathway rejoins broken double-stranded ends.
- Several different DNA polymerases that may be involved in resynthesizing stretches of replacement DNA.

Direct repair is rare and involves the reversal or simple removal of the damage.

Photoreactivation of pyrimidine dimers, in which the offending covalent bonds are reversed by a light-dependent enzyme, is a good example. This system is widespread in nature, and appears to be especially important in plants. In *E. coli* it depends on the product of a single gene (*phr*) that codes for an enzyme called photolyase.

Mismatches between the strands of DNA are one of the major targets for repair systems. **Mismatch repair** is accomplished by scrutinizing DNA for apposed bases that do not pair properly. Mismatches that arise during replication are corrected by distinguishing between the "new" and "old" strands and preferentially correcting the sequence of the newly synthesized strand. Mismatches also occur when hybrid DNA is created during recombination, and their correction upsets the ratio of parental alleles (see **Figure 15.18**). Other systems deal with mismatches generated by base conversions, such as the result of deamination. The importance of these systems is emphasized by the fact that cancer is caused in human populations by mutation of genes related to those involved in mismatch repair in yeast.

Mismatches are usually corrected by **excision repair**, which is initiated by a recognition enzyme that sees an actual damaged base or a change in the spatial path of DNA. There are two types of excision repair system.

- *Base excision repair* systems directly remove the damaged base and replace it in DNA. A good example is DNA uracil glycolase, which removes uracils that are mispaired with guanines (see *Molecular Biology 4.15.22 Base flipping is used by methylases and glycosylases*).
- *Nucleotide excision repair* systems excise a sequence that includes the damaged base(s); then a new stretch of DNA is synthesized to replace the excised material. **Figure 15.38** summarizes the main events in the operation of such a system. Such systems are common. Some recognize general damage to DNA. Others act upon specific types of base damage. There are often multiple excision repair systems in a single cell type.

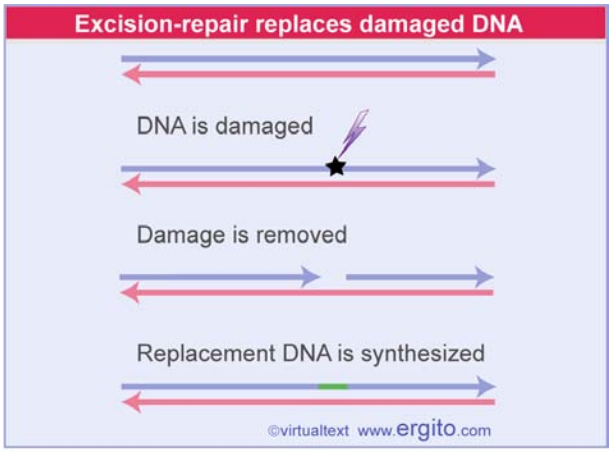

Figure 15.38 Excision-repair directly replaces damaged DNA and then resynthesizes a replacement stretch for the damaged strand.

Recombination-repair systems handle situations in which damage remains in a daughter molecule, and replication has been forced to bypass the site, typically creating a gap in the daughter strand. A retrieval system uses recombination to obtain another copy of the sequence from an undamaged source; the copy is then used to repair the gap.

A major feature in recombination and repair is the need to handle double-strand breaks. DSBs initiate crossovers in homologous recombination. They can also be created by problems in replication, when they may trigger the use of recombination-repair systems. When DSBs are created by environmental damage (for example, by radiation damage) or because of the shortening of telomeres (see *Molecular Biology 6.30.28 Genetic instability is a key event in cancer*), they can cause mutations. One system for handling DSBs can join together nonhomologous DNA ends.

Mutations that affect the ability of *E. coli* cells to engage in DNA repair fall into groups, which correspond to several repair pathways (not necessarily all independent). The major known pathways are the *uvr* excision repair system, the methyl-directed mismatch-repair system, and the *recB* and *recF* recombination and recombination-repair pathways. The enzyme activities associated with these systems are endonucleases and exonucleases (important in removing damaged DNA), resolvases (endonucleases that act specifically on recombinant junctions), helicases to unwind DNA, and DNA polymerases to synthesize new DNA. Some of these enzyme activities are unique to particular repair pathways, but others participate in multiple pathways.

The replication apparatus devotes a lot of attention to quality control. DNA polymerases use proofreading to check the daughter strand sequence and to remove errors. Some of the repair systems are less accurate when they synthesize DNA to replace damaged material. For this reason, these systems have been known historically as **error-prone** systems.

When the repair systems are eliminated, cells become exceedingly sensitive to ultraviolet irradiation. The introduction of UV-induced damage has been a major test for repair systems, and so in assessing their activities and relative efficiencies, we should remember that the emphasis might be different if another damaged adduct were studied.

Last updated on 9-16-2002

Reviews

2398. Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. (2001). *Human DNA repair genes.* Science 291, 1284-1289.

RECOMBINATION AND REPAIR

4.15.21 Excision repair systems in *E. coli*

Key Terms

Incision is a step in a mismatch excision repair system. An endonuclease recognizes the damaged area in the DNA, and isolates it by cutting the DNA strand on both sides of the damage.

The **excision** step in an excision-repair system consists of removing a single-stranded stretch of DNA by the action of a 5′-3′ exonuclease.

Key Concepts

- The Uvr system makes incisions ~12 bases apart on both sides of damaged DNA, removes the DNA between them, and resynthesizes new DNA.

Excision repair systems vary in their specificity, but share the same general features. Each system removes mispaired or damaged bases from DNA and then synthesizes a new stretch of DNA to replace them. The main type of pathway for excision repair is illustrated in **Figure 15.39**.

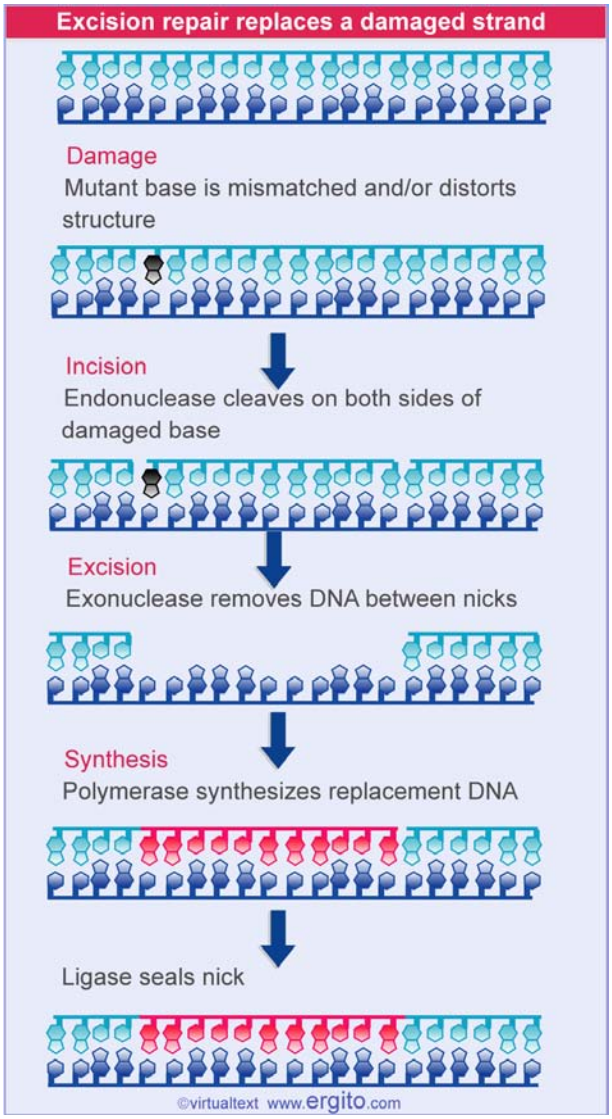

Figure 15.39 Excision-repair removes and replaces a stretch of DNA that includes the damaged base(s).

In the **incision** step, the damaged structure is recognized by an endonuclease that cleaves the DNA strand on both sides of the damage.

In the **excision** step, a 5′–3′ exonuclease removes a stretch of the damaged strand.

In the *synthesis* step, the resulting single-stranded region serves as a template for a DNA polymerase to synthesize a replacement for the excised sequence. (Synthesis of the new strand could be associated with removal of the old strand, in one coordinated action.) Finally, DNA ligase covalently links the 3′ end of the new material to the old material.

The *uvr* system of excision repair includes three genes, *uvrA,B,C*, that code for the components of a repair endonuclease. It functions in the stages indicated in **Figure 15.40**. First, a UvrAB combination recognizes pyrimidine dimers and other bulky

lesions. Then UvrA dissociates (this requires ATP), and UvrC joins UvrB. The UvrBC combination makes an incision on each side, one 7 nucleotides from the 5′ side of the damaged site, and the other 3-4 nucleotides away from the 3′ side. This also requires ATP. UvrD is a helicase that helps to unwind the DNA to allow release of the single strand between the two cuts. The enzyme that excises the damaged strand is DNA polymerase I. The enzyme involved in the repair synthesis probably also is DNA polymerase I (although DNA polymerases II and III can substitute for it). The events are basically the same, although their order is different, in the eukaryotic repair pathway shown in **Figure 15.53**.

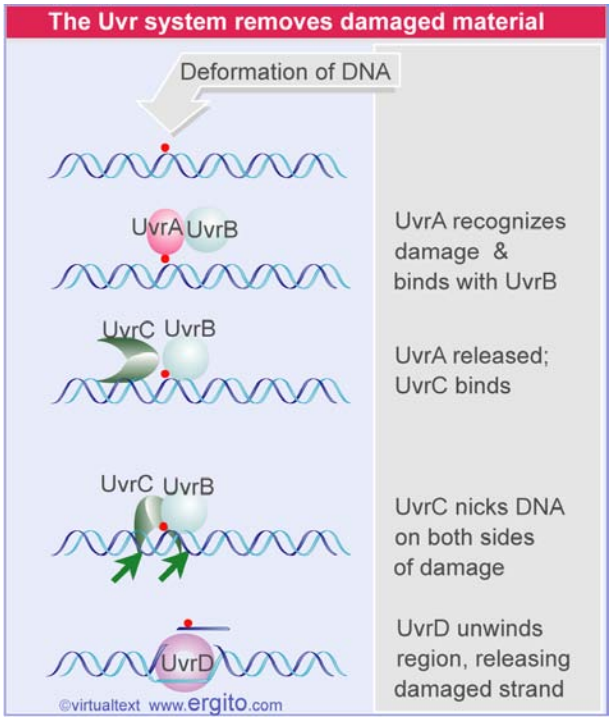

Figure 15.40 The Uvr system operates in stages in which UvrAB recognizes damage, UvrBC nicks the DNA, and UvrD unwinds the marked region.
This is a static version of an interactive figure; see http://www.ergito.com/main.jsp?bcs=MBIO.4.15.21 to view properly.

UvrABC repair accounts for virtually all of the excision repair events in *E. coli*. In almost all (99%) of cases, the average length of replaced DNA is ~12 nucleotides. (For this reason, this is sometimes described as short-patch repair). The remaining 1% involve the replacement of stretches of DNA mostly ~1500 nucleotides long, but extending up to >9000 nucleotides (sometimes called long-patch repair). We do not know why some events trigger the long-patch rather than short-patch mode.

The Uvr complex can be directed to damaged sites by other proteins. Damage to DNA may prevent transcription, but the situation is handled by a protein called Mfd that displaces the RNA polymerase and recruits the Uvr complex (see **Figure 21.19** in *Molecular Biology 5.21.12 A connection between transcription and repair*).

Last updated on 9-12-2002

RECOMBINATION AND REPAIR

4.15.22 Base flipping is used by methylases and glycosylases

Key Concepts

- Uracil and alkylated bases are recognized by glycosylases and removed directly from DNA.
- Pyrimidine dimers are reversed by breaking the covalent bonds between them.
- Methylases add a methyl group to cytosine.
- All these types of enzyme act by flipping the base out of the double helix, where, depending on the reaction, it is either removed or is modified and returned to the helix.

As an alternative to the conventional removal of part of a polynucleotide chain by nuclease activity, glycosylases and lyases can remove bases from the chain. **Figure 15.41** shows that a glycosylase cleaves the bond between the damaged or mismatched base and the deoxyribose. **Figure 15.42** shows that some glycosylases are also lyases that can take the reaction a stage further by using an amino (NH_2) group to attack the deoxyribose ring. Although the results of the glycosylase and lyase reaction appear different, the basic mechanisms of their attack on the DNA are similar (for review see 987).

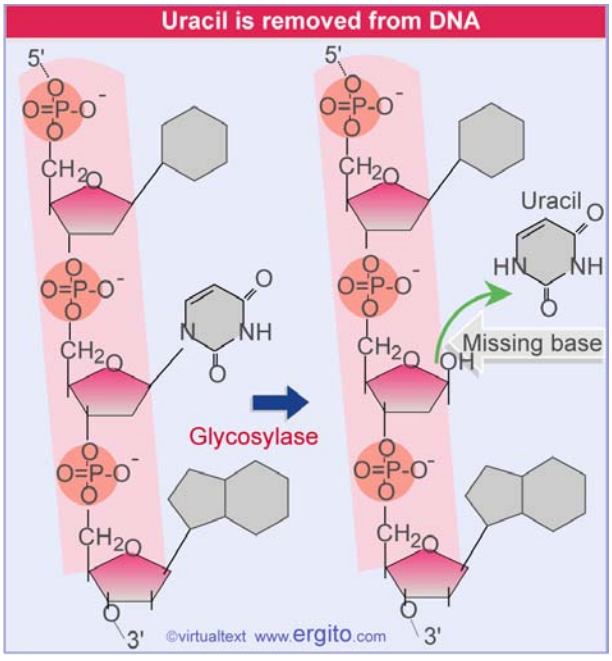

Figure 15.41 A glycosylase removes a base from DNA by cleaving the bond to the deoxyribose.

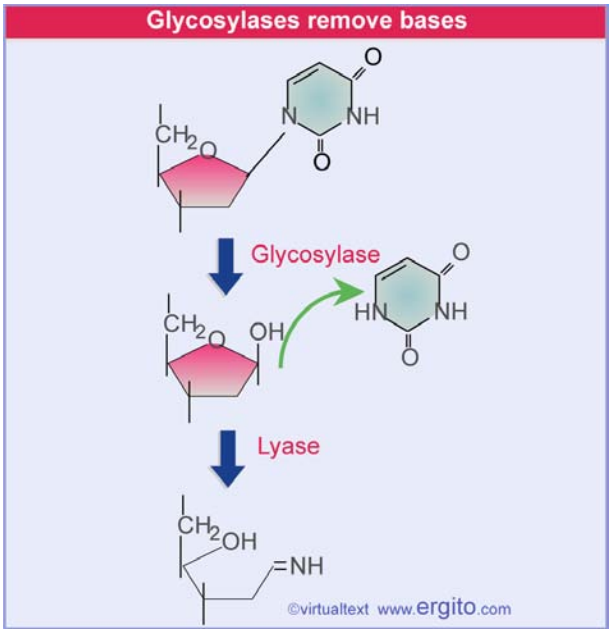

Figure 15.42 A glycosylase hydrolyzes the bond between base and deoxyribose (using H_20), but a lyase takes the reaction further by opening the sugar ring (using NH_2).

The interaction of these enzymes with DNA is remarkable. It follows the model first demonstrated for methyltransferases – enzymes that add a methyl group to cytosine in DNA. The methylase flips the target cytosine completely out of the helix (988). **Figure 15.43** shows that it enters a cavity in the enzyme where it is modified. Then it is returned to its normal position in the helix. All this occurs without input of an external energy source.

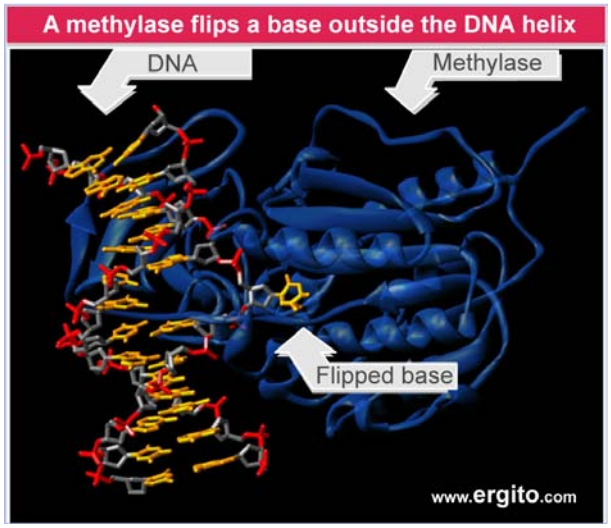

Figure 15.43 A methylase "flips" the target cytosine out of the double helix in order to modify it. Photograph kindly provided by Rich Roberts.

One of the most common reactions in which a base is directly removed from DNA is

catalyzed by uracil-DNA glycosylase. Uracil occurs in DNA most typically because of a (spontaneous) deamination of cytosine. It is recognized by the glycosylase and removed. The reaction is similar to that of the methylase: the uracil is flipped out of the helix and into the active site in the glycosylase (990; 991).

Alkylated bases (typically in which a methyl group has been added to a base) are removed by a similar mechanism (2937; 2938). A single human enzyme, alkyladenine DNA glycosylase (AAG) recognizes and removes a variety of alkylated substrates, including 3-methyladenine, 7-methylguanine, and hypoxanthine.

By contrast with this mechanism, 1-methyl-adenine is corrected by an enzyme that uses an oxygenating mechanism (coded in *E. coli* by the gene *alkB* which has homologues widely spread through nature, including three genes in Man). The methyl group is oxidized to a CH_2OH group, and then the release of the HCHO moiety [formaldehyde] restores the structure of adenine (2934; 2935). A very interesting development is the discovery that the bacterial enzyme, and one of the human enzymes, can also repair the same damaged base in RNA (3644). In the case of the human enzyme, the main target may be ribosomal RNA. This is the first known repair event with RNA as a target.

Another enzyme to use base flipping is the photolyase that reverses the bonds between pyrimidine dimers (see **Figure 15.34**). The pyrimidine dimer is flipped into a cavity in the enzyme (2939). Close to this cavity is an active site that contains an electron donor, which provides the electrons to break the bonds. Energy for the reaction is provided by light in the visible wavelength.

The common feature of these enzymes is the flipping of the target base into the enzyme structure. A variation on this theme is used by T4 endonuclease V, now renamed T4-pdg (pyrimidine dimer glycosylase) to reflect its mode of action. It flips out the adenine base that is *complementary* to the thymine on the 5′ side of the pyrimidine dimer (989). So in this case, the target for the catalytic action of the enzyme remains in the DNA duplex, and the enzyme uses flipping as an indirect mechanism to get access to its target.

When a base is removed from DNA, the reaction is followed by excision of the phosphodiester backbone by an endonuclease, DNA synthesis by a DNA polymerase to fill the gap, and ligation by a ligase to restore the integrity of the polynucleotide chain.

Last updated on 9-16-2002

Reviews

987. McCullough, A. K., Dodson, M. L., and Lloyd, R. S. (1999). *Initiation of base excision repair: glycosylase mechanisms and structures.* Annu. Rev. Biochem. 68, 255-285.

References

988. Klimasauskas, S., Kumar, S., Roberts, R. J. and Cheng, X. (1994). *HhaI methyltransferase flips its target base out of the DNA helix.* Cell 76, 357-369.

989. Vassylyev, D. G. et al. (1995). *Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition.* Cell 83, 773-782.

990. Savva, R. et al. (1995). *The structural basis of specific base-excision repair by uracil-DNA glycosylase.* Nature 373, 487-493.

991. Mol, D. D. et al. (1995). *Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis.* Cell 80, 869-878.

2934. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., and Sedgwick, B. (2002). *Oxidative demethylation by E. coli AlkB directly reverts DNA base damage.* Nature 419, 174-178.

2935. Falnes, P. A., Johansen, R. F., and Seeberg, E. (2002). *AlkB-mediated oxidative demethylation reverses DNA damage in E. coli.* Nature 419, 178-182.

2937. Lau, A. Y., Scherer, O. D., Samson, L., Verdine, G. L., and Ellenberger, T. (1998). *Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision.* Cell 95, 249-258.

2938. Lau, A. Y., Glassner, B. J., Samson, L. D., and Ellenberger, T. (2000). *Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG.* Proc. Natl. Acad. Sci. USA 97, 13573-13578.

2939. Park, H. W., Kim, S. T., Sancar, A., and Deisenhofer, J. (1995). *Crystal structure of DNA photolyase from E. coli.* Science 268, 1866-1872.

3644. Aas, P. A., Otterlei, M., Falnes, P. A., Vagbe, C. B., Skorpen, F., Akbari, M., Sundheim, O., Bjoras, M., Slupphaug, G., Seeberg, E., and Krokan, H. E. (2003). *Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA.* Nature 421, 859-863.

RECOMBINATION AND REPAIR

4.15.23 Error-prone repair and mutator phenotypes

Key Terms

Error-prone synthesis occurs when DNA incorporates noncomplementary bases into the daughter strand.

Key Concepts

- Damaged DNA that has not been repaired causes DNA polymerase III to stall during replication.
- DNA polymerase V (coded by *umuCD*), or DNA polymerase IV (coded by *dinB*) can synthesize a complement to the damaged strand.
- The DNA synthesized by the repair DNA polymerase often has errors in its sequence.
- Proteins that affect the fidelity of replication may be identified by mutator genes, in which mutation causes an increased rate of spontaneous mutation.

The existence of repair systems that engage in DNA synthesis raises the question of whether their quality control is comparable with that of DNA replication. So far as we know, most systems, including *uvr*-controlled excision repair, do not differ significantly from DNA replication in the frequency of mistakes. However, **error-prone** synthesis of DNA occurs in *E. coli* under certain circumstances.

The error-prone feature was first observed when it was found that the repair of damaged λ phage DNA is accompanied by the induction of mutations if the phage is introduced into cells that had previously been irradiated with UV. This suggests that the UV irradiation of the host has activated functions that generate mutations. The mutagenic response also operates on the bacterial host DNA.

What is the actual error-prone activity? It is a DNA polymerase that inserts incorrect bases, which represent mutations, when it passes any site at which it cannot insert complementary base pairs in the daughter strand. Functions involved in this error-prone pathway are identified by mutations in the genes *umuD* and *umuC*, which abolish UV-induced mutagenesis. This implies that the UmuC and UmuD proteins cause mutations to occur after UV irradiation. The genes constitute the *umuDC* operon, whose expression is induced by DNA damage. Their products form a complex UmuD′$_2$C, consisting of two subunits of a truncated UmuD protein and one subunit of UmuC. UmuD is cleaved by RecA, which is activated by DNA damage (see *Molecular Biology 4.15.27 RecA triggers the SOS system*).

The UmuD′$_2$C complex has DNA polymerase activity. It is called DNA polymerase V, and is responsible for synthesizing new DNA to replace sequences that have been damaged by UV (1159). This is the only enzyme in *E. coli* that can bypass the classic pyrimidine dimers produced by UV (or other bulky adducts). The polymerase

activity is "error-prone". Mutation of *umuC* or *umuD* inactivate the enzyme, which makes UV irradiation lethal. Some plasmids carry genes called *mucA* and *mucB*, which are homologues of *umuD* and *umuC*, and whose introduction into a bacterium increases resistance to UV killing and susceptibility to mutagenesis (1564).

How does an alternative DNA polymerase get access to the DNA? When the replicase (DNA polymerase III) encounters a block, such as a thymidine dimer, it stalls. Then it is displaced from the replication fork and replaced by DNA polymerase V. In fact, DNA polymerase V uses some of the same ancillary proteins as DNA polymerase III. The same situation is true for DNA polymerase IV, the product of *dinB*, which is another enzyme that acts on damaged DNA (1166). DNA polymerases IV and V are part of a larger family, including eukaryotic DNA polymerases, that are involved in repairing damaged DNA (see *Molecular Biology 4.15.28 Eukaryotic cells have conserved repair systems*) (for review see 1164).

Last updated on 9-27-2000

References

1159. Maor-Shoshani, A., Reuven, N. B., Tomer, G., and Livneh, Z. (2000). *Highly mutagenic replication by DNA polymerase V (UmuC) provides a mechanistic basis for SOS untargeted mutagenesis.* Proc. Natl. Acad. Sci. USA 97, 565-570.

1164. Friedberg, E. C., Feaver, W. J., and Gerlach, V. L. (2000). *The many faces of DNA polymerases: strategies for mutagenesis and for mutational avoidance*. Proc. Natl. Acad. Sci. USA 97, 5681-5683.

1166. Wagner, J., Gruz, P., Kim, S. R., Yamada, M., Matsui, K., Fuchs, R. P., and Nohmi, T. (1999). *The dinB gene encodes a novel E. coli DNA polymerase, DNA pol IV, involved in mutagenesis.* Mol. Cell 4, 281-286.

1564. Goldsmith, M., Sarov-Blat, L., and Livneh, Z. (2000). *Plasmid-encoded MucB protein is a DNA polymerase (pol RI) specialized for lesion bypass in the presence of MucA, RecA, and SSB.* Proc. Natl. Acad. Sci. USA 97, 11227-11231.

RECOMBINATION AND REPAIR

4.15.24 Controlling the direction of mismatch repair

Key Terms

A **mutator** is a mutation or a mutated gene that increases the basal level of mutation. Such genes often code for proteins that are involved in repairing damaged DNA.

Key Concepts

- The *mut* genes code for a mismatch repair system that deals with mismatched base pairs.
- There is a bias in the selection of which strand to replace at mismatches.
- The strand lacking methylation at a hemimethylated is usually replaced.
- This is used to remove errors in a newly synthesized strand of DNA. At G·T and C·T mismatches, the T is preferentially removed.

Genes whose products are involved in controlling the fidelity of DNA synthesis during either replication or repair may be identified by mutations that have a **mutator** phenotype. A mutator mutant has an increased frequency of spontaneous mutation. If identified originally by the mutator phenotype, a gene is described as *mut*; but often a *mut* gene is later found to be equivalent with a known replication or repair activity.

The general types of activities are identified by *mut* genes fall into two groups.

- The major group consists of components of mismatch-repair systems. Failure to remove a damaged or mispaired base before replication allows it to induce a mutation. Functions in this group include the *dam* methylase that identifies the target for repair, and enzymes that participate directly or indirectly in the removal of particular types of damage (*mutH,S,L,Y).*
- A smaller group, typified by *dnaQ* (which codes for a subunit of DNA polymerase III), is concerned with the accuracy of synthesizing new DNA.

When a structural distortion is removed from DNA, the wild-type sequence is restored. In most cases, the distortion is due to the creation of a base that is not naturally found in DNA, and which is therefore recognized and removed by the repair system.

A problem arises if the target for repair is a mispaired partnership of (normal) bases created when one was mutated. The repair system has no intrinsic means of knowing which is the wild-type base and which is the mutant! All it sees are two improperly paired bases, either of which can provide the target for excision repair.

If the mutated base is excised, the wild-type sequence is restored. But if it happens to be the original (wild-type) base that is excised, the new (mutant) sequence becomes fixed. Often, however, the direction of excision repair is not random, but is biased in a way that is likely to lead to restoration of the wild-type sequence.

Some precautions are taken to direct repair in the right direction. For example, for cases such as the deamination of 5-methyl-cytosine to thymine, there is a special system to restore the proper sequence (see also *Molecular Biology 1.1.14 Many hotspots result from modified bases*). The deamination generates a G·T pair, and the system that acts on such pairs has a bias to correct them to G·C pairs (rather than to A·T pairs). The system that undertakes this reaction includes the *mutL,S* products that remove T from both G·T and C·T mismatches.

The *mutT,M,Y* system handles the consequences of oxidative damage. A major type of chemical damage is caused by oxidation of G to 8-oxo-G. **Figure 15.44** shows that the system operates at three levels. MutT hydrolyzes the damaged precursor (8-oxo-dGTP), which prevents it from being incorporated into DNA. When guanine is oxidized in DNA, its partner is cytosine; and MutM preferentially removes the C from 8-oxo-G·C pairs. Oxidized guanine mispairs with A, and so when 8-oxo-G survives and is replicated, it generates an 8-oxo-G·A pair. MutY removes A from these pairs. MutM and MutY are glycosylases that directly remove a base from DNA. This creates a apurinic site that is recognized by an endonuclease whose action triggers the involvement of the excision repair system.

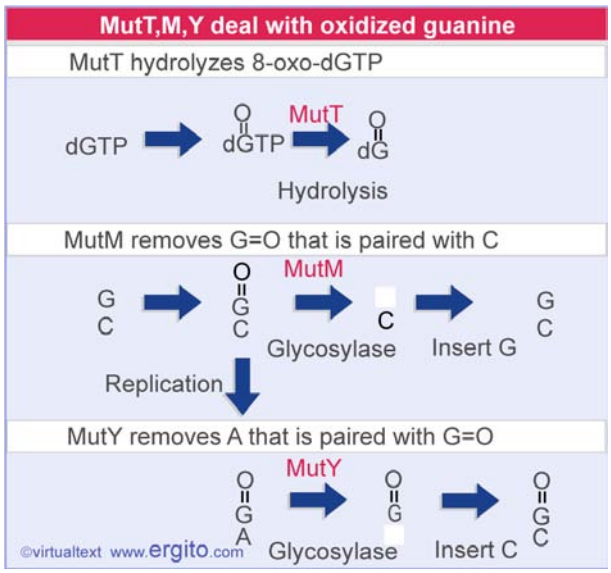

Figure 15.44 Preferential removal of bases in pairs that have oxidized guanine is designed to minimize mutations.

When mismatch errors occur during replication in *E. coli*, it is possible to distinguish the original strand of DNA. Immediately after replication of methylated DNA, only the original parental strand carries the methyl groups. In the period while the newly synthesized strand awaits the introduction of methyl groups, the two strands can be distinguished.

This provides the basis for a system to correct replication errors. The *dam* gene codes

for a methylase whose target is the adenine in the sequence (see **Figure 14.35**). The hemimethylated state is used to distinguish replicated origins from nonreplicated origins. The same target sites are used by a replication-related repair system.

Figure 15.45 shows that DNA containing mismatched base partners is repaired preferentially by excising the strand that lacks the methylation. The excision is quite extensive; mismatches can be repaired preferentially for >1 kb around a GATC site. The result is that the newly synthesized strand is corrected to the sequence of the parental strand.

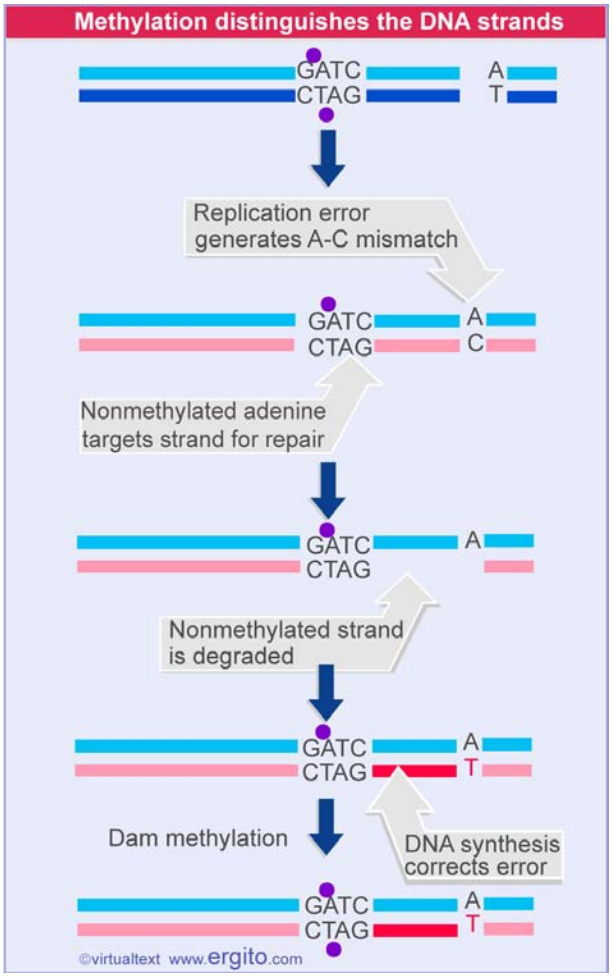

Figure 15.45 GATC sequences are targets for the Dam methylase after replication. During the period before this methylation occurs, the nonmethylated strand is the target for repair of mismatched bases.

E. coli dam$^-$ mutants show an increased rate of spontaneous mutation. This repair system therefore helps reduce the number of mutations caused by errors in replication. It consists of several proteins, coded by the *mut* genes. MutS binds to the mismatch and is joined by MutL. MutS can use two DNA-binding sites, as illustrated in **Figure 15.46**. The first specifically recognizes mismatches. The second is not specific for sequence or structure, and is used to translocate along DNA until a GATC sequence is encountered. Hydrolysis of ATP is used to drive the

translocation. Because MutS is bound to both the mismatch site and to DNA as it translocates, it creates a loop in the DNA.

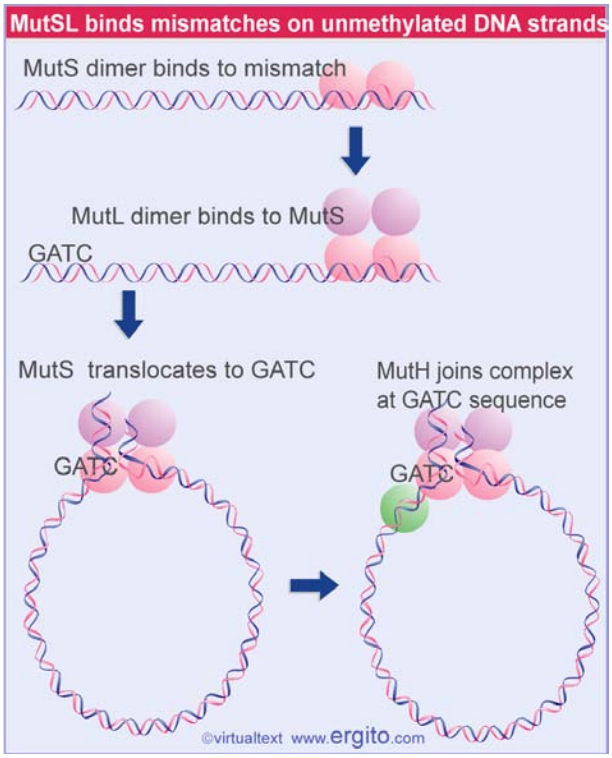

Figure 15.46 MutS recognizes a mismatch and translocates to a GATC site. MutH cleaves the unmethylated strand at the GATC. Endonucleases degrade the strand from the GATC to the mismatch site.

Recognition of the GATC sequence causes the MutH endonuclease to bind to MutSL. The endonuclease then cleaves the unmethylated strand. This strand is then excised from the GATC site to the mismatch site. The excision can occur in either the 5′–3′ direction (using RecJ or exonuclease VII) or in the 3′–5′ direction (using exonuclease I), assisted by the helicase UvrD. The new DNA strand is synthesized by DNA polymerase III.

The *msh* repair system of *S. cerevisiae* is homologous to the *E. coli mut* system. MSH2 provides a scaffold for the apparatus that recognizes mismatches. MSH3 and MSH6 provide specificity factors. The MSH2-MSH3 complex binds mismatched loops of 2-4 nucleotides, and the MSH2-MSH6 complex binds to single base mismatches or insertions or deletions. Other proteins are then required for the repair process itself.

homologues of the MutSL system also are found in higher eukaryotic cells. They are responsible for repairing mismatches that arise as the result of replication slippage (2255). In a region such as a microsatellite where a very short sequence is repeated several times, realignment between the newly synthesized daughter strand and its template can lead to a stuttering in which the DNA polymerase slips backward and synthesizes extra repeating units. These units in the daughter strand are extruded as a

single-stranded loop from the double helix (see **Figure 4.28**). They are repaired by homologues of the MutSL system as shown in **Figure 15.47**.

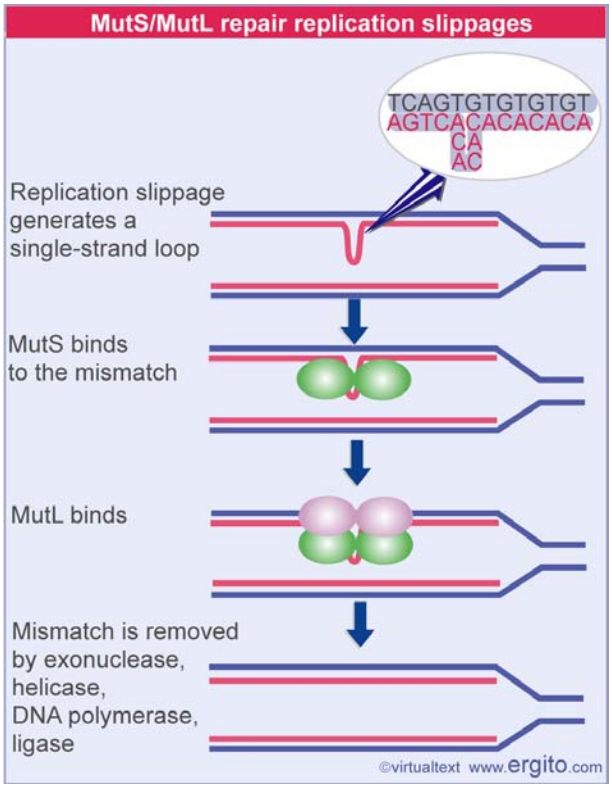

Figure 15.47 The MutS/MutL system initiates repair of mismatches produced by replication slippage.

The importance of the MutSL system for mismatch repair is indicated by the high rate at which it is found to be defective in human cancers. Loss of this system leads to an increased mutation rate (see *Molecular Biology 6.30.29 Defects in repair systems cause mutations to accumulate in tumors*).

Last updated on 1-18-2002

References

2255. Strand, M., Prolla, T. A., Liskay, and Petes, T. D. (1993). *Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair.* Nature 365, 274-276.

RECOMBINATION AND REPAIR

4.15.25 Recombination-repair systems in *E. coli*

Key Terms

Single-strand exchange is a reaction in which one of the strands of a duplex of DNA leaves its former partner and instead pairs with the complementary strand in another molecule, displacing its homologue in the second duplex.

Key Concepts

- The *rec* genes of *E. coli* code for the principal retrieval system.
- It functions when replication leaves a gap in newly synthesized strand opposite a damaged sequence.
- The single strand of another duplex is used to replace the gap.
- Then the damaged sequence is removed and resynthesized.

Recombination-repair systems use activities that overlap with those involved in genetic recombination. They are also sometimes called "post-replication repair," because they function after replication. Such systems are effective in dealing with the defects produced in daughter duplexes by replication of a template that contains damaged bases. An example is illustrated in **Figure 15.48**. Restarting stalled replication forks could be the major role of the recombination-repair systems (see *Molecular Biology 4.14.17 The primosome is needed to restart replication*).

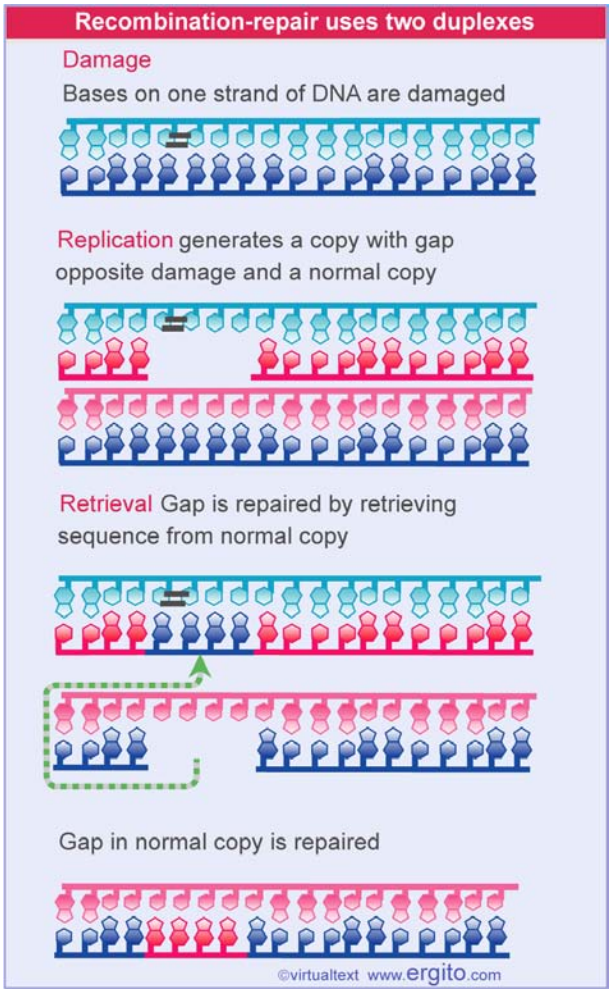

Figure 15.48 An *E. coli* retrieval system uses a normal strand of DNA to replace the gap left in a newly synthesized strand opposite a site of unrepaired damage.

Consider a structural distortion, such as a pyrimidine dimer, on one strand of a double helix. When the DNA is replicated, the dimer prevents the damaged site from acting as a template. Replication is forced to skip past it.

DNA polymerase probably proceeds up to or close to the pyrimidine dimer. Then the polymerase ceases synthesis of the corresponding daughter strand. Replication restarts some distance farther along. A substantial gap is left in the newly synthesized strand.

The resulting daughter duplexes are different in nature. One has the parental strand containing the damaged adduct, facing a newly synthesized strand with a lengthy gap. The other duplicate has the undamaged parental strand, which has been copied into a normal complementary strand. The retrieval system takes advantage of the normal daughter.

The gap opposite the damaged site in the first duplex is filled by stealing the homologous single strand of DNA from the normal duplex. Following this

single-strand exchange, the recipient duplex has a parental (damaged) strand facing a wild-type strand. The donor duplex has a normal parental strand facing a gap; the gap can be filled by repair synthesis in the usual way, generating a normal duplex. So the damage is confined to the original distortion (although the same recombination-repair events must be repeated after every replication cycle unless and until the damage is removed by an excision repair system).

The principal pathway for recombination-repair in *E. coli* is identified by the *rec* genes (see **Figure 15.13**, **Figure 15.14**, **Figure 15.15**) (for review see 3131). In *E. coli* deficient in excision repair, mutation in the *recA* gene essentially abolishes all the remaining repair and recovery facilities. Attempts to replicate DNA in $uvr^{-}recA^{--}$ cells produce fragments of DNA whose size corresponds with the expected distance between thymine dimers. This result implies that the dimers provide a lethal obstacle to replication in the absence of RecA function. It explains why the double mutant cannot tolerate >1-2 dimers in its genome (compared with the ability of a wild-type bacterium to handle as many as 50).

One *rec* pathway involves the *recBC* genes, and is well characterized; the other involves *recF*, and is not so well defined. They fulfill different functions *in vivo*. The RecBC pathway is involved in restarting stalled replication forks (see *Molecular Biology 4.15.26 Recombination is an important mechanism to recover from replication errors*). The RecF pathway is involved in repairing the gaps in a daughter strand that are left after replicating past a pyrimidine dimer.

The RecBC and RecF pathways both function prior to the action of RecA (although in different ways). They lead to the association of RecA with a single-stranded DNA. The ability of RecA to exchange single strands allows it to perform the retrieval step in **Figure 15.48**. Nuclease and polymerase activities then complete the repair action.

The RecF pathway contains a group of three genes: *recF*, *recO*, and *recR*. The proteins form two types of complex, RecOR and RecOF. They promote the formation of RecA filaments on single-stranded DNA. One of their functions it to make it possible for the filaments to assemble in spite of the presence of the SSB, which is inhibitory. Although they are thought to function at gaps, the reaction *in vitro* requires a free 5′ end (3134).

The designations of repair and recombination genes are based on the phenotypes of the mutants; but sometimes a mutation isolated in one set of conditions and named as a *uvr* locus turns out to have been isolated in another set of conditions as a *rec* locus. This uncertainty makes an important point. We cannot yet define how many functions belong to each pathway or how the pathways interact. The *uvr* and *rec* pathways are not entirely independent, because *uvr* mutants show reduced efficiency in recombination-repair. We must expect to find a network of nuclease, polymerase, and other activities, constituting repair systems that are partially overlapping (or in which an enzyme usually used to provide some function can be substituted by another from a different pathway).

Last updated on 11-30-2002

Reviews

3131. West, S. C. (1997). *Processing of recombination intermediates by the RuvABC proteins.* Annu. Rev. Genet. 31, 213-244.

References

3134. Bork, J. M. and Inman, R. B. (2001). *The RecOR proteins modulate RecA protein function at 5'' ends of single-stranded DNA.* EMBO J. 20, 7313-7322.

RECOMBINATION AND REPAIR

4.15.26 Recombination is an important mechanism to recover from replication errors

Key Concepts

- A replication fork may stall when it encounters a damaged site or a nick in DNA.
- A stalled fork may reverse by pairing between the two newly synthesized strands.
- A stalled fork may restart repairing the damage and using a helicase to move the fork forward.
- The structure of the stalled fork is the same as a Holliday junction and may be converted to a duplex and DSB by resolvases.

All cells have many pathways to repair damage in DNA. Which pathway is used will depend upon the type of damage and the situation. Excision-repair pathways can in principle be used at any time, but recombination-repair can be used only when there is a second duplex with a copy of the damaged sequence, that is, post-replication. A special situation is presented when damaged DNA is replicated, because the replication fork may stall at the site of damage. Recombination-repair pathways are involved in allowing the fork to be restored after the damage has been repaired or to allow it to bypass the damage.

Figure 15.49 shows one possible outcome when a replication fork stalls (for review see 956; 3442). The fork stops moving forward when it encounters the damage. The replication apparatus disassembles, at least partially. This allows branch migration to occur, when the fork effectively moves backward, and the new daughter strands pair to form a duplex structure. After the damage has been repaired, a helicase rolls the fork forward to restore its structure. Then the replication apparatus can reassemble, and replication is restarted (see *Molecular Biology 4.14.17 The primosome is needed to restart replication*). DNA polymerase II is required for the replication restart, and is later replaced by DNA polymerase III (4524).

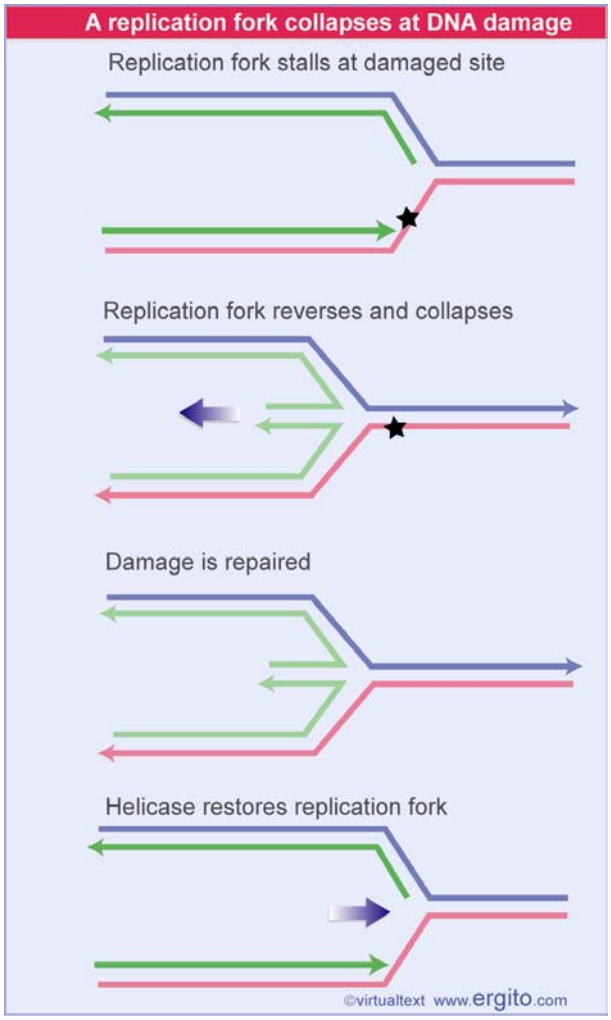

Figure 15.49 A replication fork stalls when it reaches a damaged site in DNA. Branch migration occurs to move the fork backward, and the two daughter strands pair to form a duplex. After the damage has been repaired, a helicase may cause forward branch migration to restore the structure of the fork. Arrowheads indicate 3′ ends.

The pathway for handling a stalled replication fork requires repair enzymes. In *E. coli*, RecA and the RecBC system have an important role in this reaction (in fact, this may be their major function in the bacterium). One possible pathway is for RecA to bind to single-stranded DNA at the stalled replication fork, stabilizing it, and possibly acting as the sensor that detects the stalling event. RecBC is involved in excision-repair of the damage. After the damage has been repaired, replication can resume (for review see 4525).

Another pathway may use recombination-repair, possibly the strand-exchange reactions of RecA. **Figure 15.50**shows that the structure of the stalled fork is essentially the same as a Holliday junction created by recombination between two duplex DNAs. This makes it a target for resolvases. A double-strand break is generated if a resolvase cleaves either pair of complementary strands. In addition, if the damage is in fact a nick, another double-strand break is created at this site (2928).

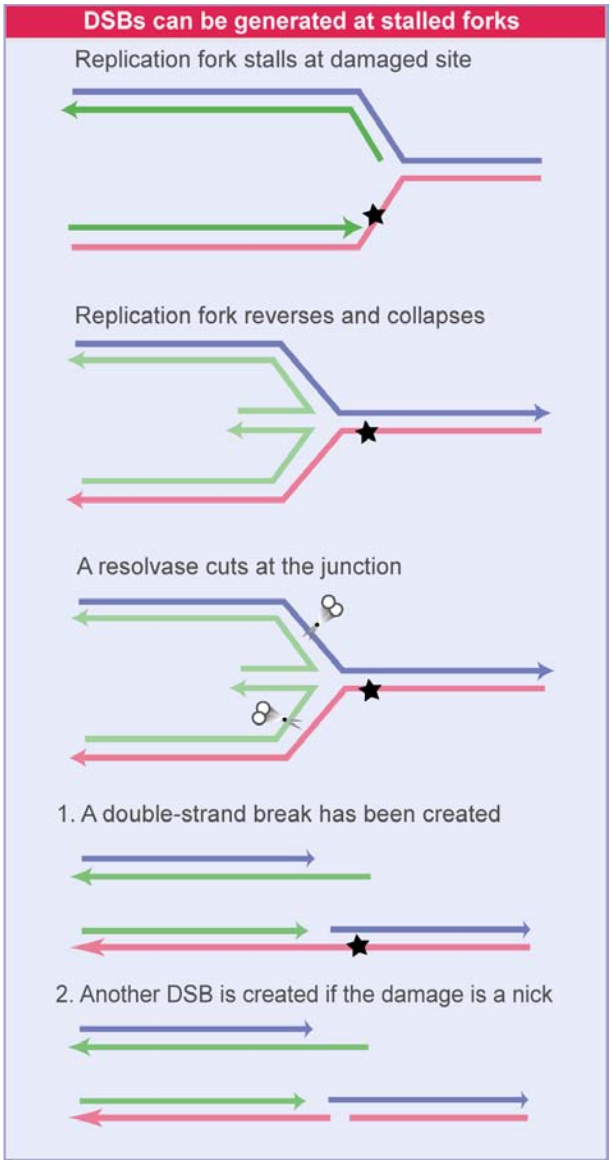

Figure 15.50 The structure of a stalled replication fork resembles a Holliday junction and can be resolved in the same way by resolvases. The results depend on whether the site of damage contains a nick. Result 1 shows that a double-strand break is generated by cutting a pair of strands at the junction. Result 2 shows a second DSB is generated at the site of damage if it contains a nick. Arrowheads indicate 3′ ends.

Stalled replication forks can be rescued by recombination-repair (for review see 2929). We don't know the exact sequence of events, but one possible scenario is outlined in **Figure 15.51**. The principle is that a recombination event occurs on either side of the damaged site, allowing an undamaged single strand to pair with the damaged strand. This allows the replication fork to be reconstructed, so that replication can continue, effectively bypassing the damaged site.

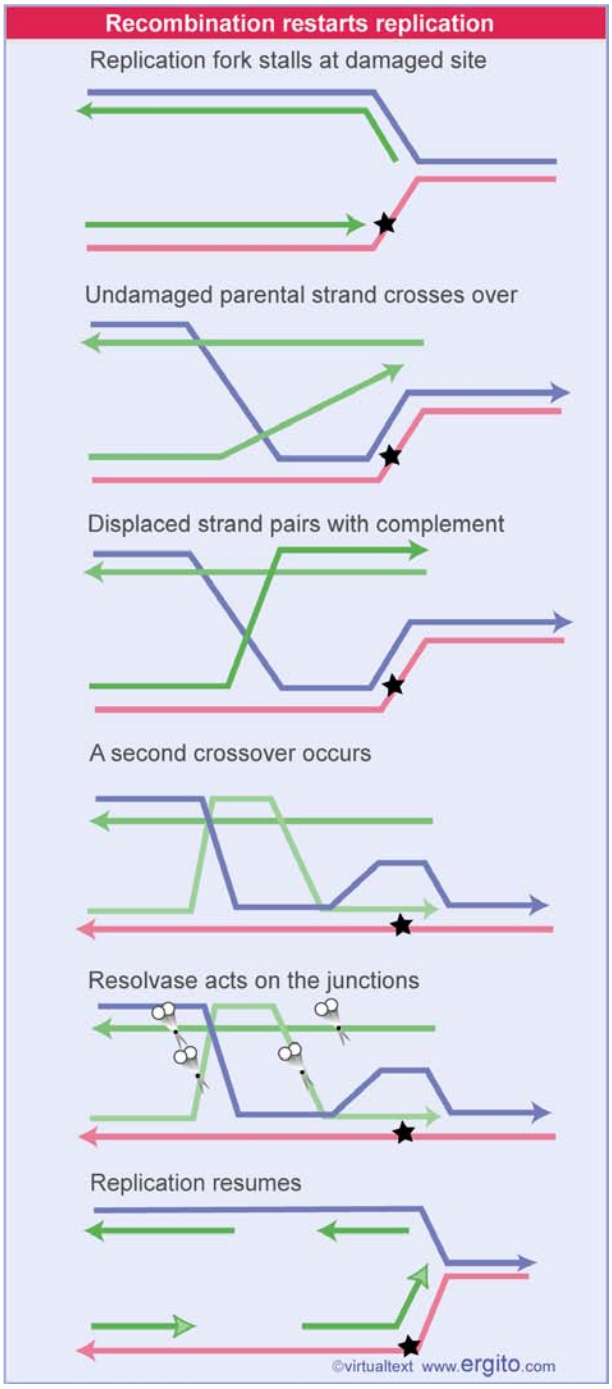

Figure 15.51 When a replication fork stalls, recombination-repair can place an undamaged strand opposite the damaged site. This allows replication to continue.

Last updated on January 9, 2004

Reviews

2929. Michel, B., Viguera, E., Grompone, G., Seigneur, M., and Bidnenko, V. (2001). *Rescue of arrested replication forks by homologous recombination.* Proc. Natl. Acad. Sci. USA 98, 8181-8188.

3442. McGlynn, P. and Lloyd, R. G. (2002). *Recombinational repair and restart of damaged replication forks.* Nat. Rev. Mol. Cell Biol. 3, 859-870.

References

956. Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J., and Marians, K. J. (2000). *The importance of repairing stalled replication forks.* Nature 404, 37-41.

2928. Kuzminov, A. (2001). *Single-strand interruptions in replicating chromosomes cause double-strand breaks.* Proc. Natl. Acad. Sci. USA 98, 8241-8246.

4524. Rangarajan, S., Woodgate, R., and Goodman, M. F. (1999). *A phenotype for enigmatic DNA polymerase II: a pivotal role for pol II in replication restart in UV-irradiated Escherichia coli.* Proc. Natl. Acad. Sci. USA 96, 9224-9229.

4525. Courcelle, J. and Hanawalt, P. C. (2003). *RecA-dependent recovery of arrested DNA replication forks.* Annu. Rev. Genet. 37, 611-646.

RECOMBINATION AND REPAIR

4.15.27 RecA triggers the SOS system

Key Terms

An **SOS response** in *E. coli* describes the coordinate induction of many enzymes, including repair activities, in response to irradiation or other damage to DNA; results from activation of protease activity by RecA to cleave LexA repressor.

The **SOS box** is the DNA sequence (operator) of ~20 bp recognized by LexA repressor protein.

Key Concepts

- Damage to DNA causes RecA to trigger the SOS response, consisting of genes coding for many repair enzymes.
- RecA activates the autocleavage activity of LexA.
- LexA represses the SOS system; its autocleavage activates those genes.

The direct involvement of RecA protein in recombination-repair is only one of its activities. This extraordinary protein also has another, quite distinct function. It can be activated by many treatments that damage DNA or inhibit replication in *E. coli*. This causes it to trigger a complex series of phenotypic changes called the **SOS response**, which involves the expression of many genes whose products include repair functions. These dual activities of the RecA protein make it difficult to know whether a deficiency in repair in *recA* mutant cells is due to loss of the DNA strand-exchange function of RecA or to some other function whose induction depends on the protease activity.

The inducing damage can take the form of ultraviolet irradiation (the most studied case) or can be caused by crosslinking or alkylating agents. Inhibition of replication by any of several means, including deprivation of thymine, addition of drugs, or mutations in several of the *dna* genes, has the same effect.

The response takes the form of increased capacity to repair damaged DNA, achieved by inducing synthesis of the components of both the long-patch excision repair system and the Rec recombination-repair pathways. In addition, cell division is inhibited. Lysogenic prophages may be induced.

The initial event in the response is the activation of RecA by the damaging treatment. We do not know very much about the relationship between the damaging event and the sudden change in RecA activity. Because a variety of damaging events can induce the SOS response, current work focuses on the idea that RecA is activated by some common intermediate in DNA metabolism.

The inducing signal could consist of a small molecule released from DNA; or it might be some structure formed in the DNA itself. *In vitro*, the activation of RecA

requires the presence of single-stranded DNA and ATP. So the activating signal could be the presence of a single-stranded region at a site of damage. Whatever form the signal takes, its interaction with RecA is rapid: the SOS response occurs within a few minutes of the damaging treatment.

Activation of RecA causes proteolytic cleavage of the product of the *lexA* gene. LexA is a small (22 kD) protein that is relatively stable in untreated cells, where it functions as a repressor at many operons. The cleavage reaction is unusual; LexA has a latent protease activity that is activated by RecA. When RecA is activated, it causes LexA to undertake an autocatalytic cleavage; this inactivates the LexA repressor function, and coordinately induces all the operons to which it was bound. The pathway is illustrated in **Figure 15.52**.

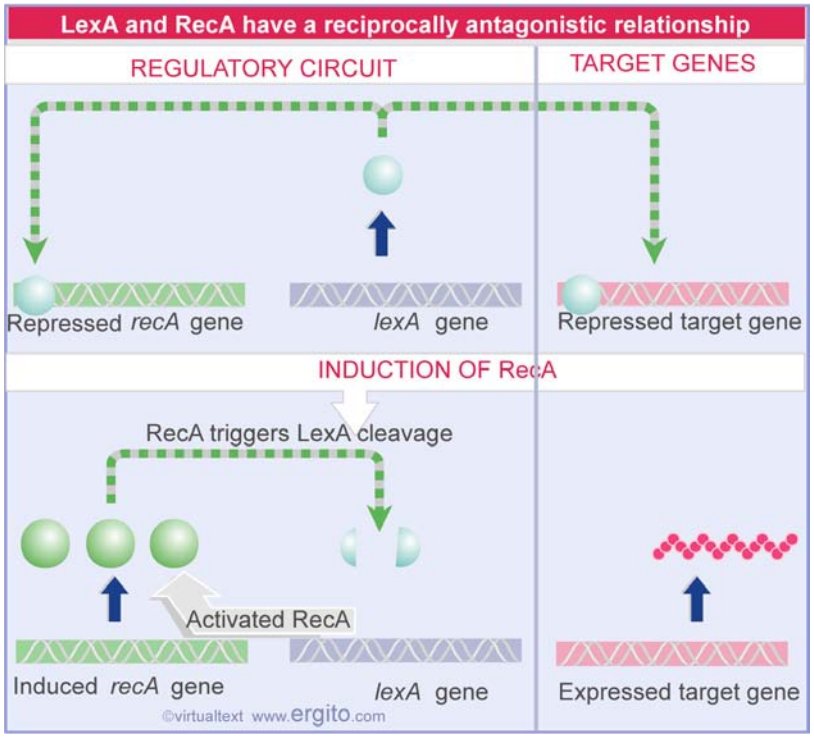

Figure 15.52 The LexA protein represses many genes, including repair functions, *recA* and *lexA*. Activation of RecA leads to proteolytic cleavage of LexA and induces all of these genes.

The target genes for LexA repression include many repair functions. Some of these SOS genes are active only in treated cells; others are active in untreated cells, but the level of expression is increased by cleavage of LexA. In the case of *uvrB*, which is a component of the excision repair system, the gene has two promoters; one functions independently of LexA, the other is subject to its control. So after cleavage of LexA, the gene can be expressed from the second promoter as well as from the first.

LexA represses its target genes by binding to a 20 bp stretch of DNA called an **SOS box**, which includes a consensus sequence with 8 absolutely conserved positions. Like other operators, the SOS boxes overlap with the respective promoters. At the *lexA* locus, the subject of autogenous repression, there are two adjacent SOS boxes.

RecA and LexA are mutual targets in the SOS circuit: RecA triggers cleavage of LexA, which represses *recA* and itself. The SOS response therefore causes amplification of both the RecA protein and the LexA repressor. The results are not so contradictory as might at first appear.

The increase in expression of RecA protein is necessary (presumably) for its direct role in the recombination-repair pathways. On induction, the level of RecA is increased from its basal level of ~1200 molecules/cell by up to 50×. The high level in induced cells means there is sufficient RecA to ensure that all the LexA protein is cleaved. This should prevent LexA from reestablishing repression of the target genes.

But the main importance of this circuit for the cell lies in the ability to return rapidly to normalcy. When the inducing signal is removed, the RecA protein loses the ability to destabilize LexA. At this moment, the *lexA* gene is being expressed at a high level; in the absence of activated RecA, the LexA protein rapidly accumulates in the uncleaved form and turns off the SOS genes. This explains why the SOS response is freely reversible.

RecA also triggers cleavage of other cellular targets, sometimes with more direct consequences. The UmuD protein is cleaved when RecA is activated; the cleavage event activates UmuD and the error-prone repair system. The current model for the reaction is that the $UmuD_2$UmuC complex binds to a RecA filament near a site of damage, RecA activates the complex by cleaving UmuD to generate UmuD′, and the complex then synthesizes a stretch of DNA to replace the damaged material (943).

Activation of RecA also causes cleavage of some other repressor proteins, including those of several prophages. Among these is the lambda repressor (with which the protease activity was discovered). This explains why lambda is induced by ultraviolet irradiation; the lysogenic repressor is cleaved, releasing the phage to enter the lytic cycle.

This reaction is not a cellular SOS response, but instead represents a recognition by the prophage that the cell is in trouble. Survival is then best assured by entering the lytic cycle to generate progeny phages. In this sense, prophage induction is piggybacking onto the cellular system by responding to the same indicator (activation of RecA).

The two activities of RecA are relatively independent. The *recA441* mutation allows the SOS response to occur without inducing treatment, probably because RecA remains spontaneously in the activated state. Other mutations abolish the ability to be activated. Neither type of mutation affects the ability of RecA to handle DNA. The reverse type of mutation, inactivating the recombination function but leaving intact the ability to induce the SOS response, would be useful in disentangling the direct and indirect effects of RecA in the repair pathways.

Last updated on 1-27-2000

References

943. Tang, M. et al. (1999). *UmuD'$_2$C is an error-prone DNA polymerase, E. coli pol V.* Proc. Natl. Acad. Sci. USA 96, 8919-8924.

RECOMBINATION AND REPAIR

4.15.28 Eukaryotic cells have conserved repair systems

Key Concepts

- The yeast *RAD* mutations, identified by radiation sensitive phenotypes, are in genes that code for repair systems.
- Xeroderma pigmentosum (XP) is a human disease caused by mutations in any one of several repair genes.
- A complex of proteins including XP products and the transcription factor TF_{IIH} provides a human excision-repair mechanism.
- Transcriptionally active genes are preferentially repaired.

The types of repair functions recognized in *E. coli* are common to a wide range of organisms. The best characterized eukaryotic systems are in yeast, where Rad51 is the counterpart to RecA. In yeast, the main function of the strand-transfer protein is homologous recombination. Many of the repair systems found in yeast have direct counterparts in higher eukaryotic cells, and in several cases these systems are involved with human diseases (see also *Molecular Biology 6.30.29 Defects in repair systems cause mutations to accumulate in tumors*).

Genes involved in repair functions have been characterized genetically in yeast by virtue of their sensitivity to radiation. They are called *RAD* genes. There are three general groups of repair genes in the yeast *S. cerevisiae*, identified by the *RAD3* group (involved in excision repair), the *RAD6* group (required for post-replication repair), and the *RAD52* group (concerned with recombination-like mechanisms). The RAD52 group is divided into two subgroups by a difference in mutant phenotypes. One subgroup affects homologous recombination, as seen by a reduction in mitotic recombination in *RAD50*, *RAD51*, *RAD54*, *RAD55*, and *RAD57*. These Rad proteins form a multiprotein complex at a double-strand break. After an exonuclease has acted on the free ends to generate single-stranded tails, Rad51 initiates the process by binding to the single-stranded DNA to form a nucleoprotein filament. Rad52, Rad55, and Rad54 then bind sequentially to the filament (4190). By contrast, recombination rates are increased in *RAD59*, *MRE11*, and *XRS2* mutants; this subgroup is not deficient in homologous recombination, but is deficient in nonhomologous DNA joining reactions.

A superfamily of DNA polymerases involved in synthesizing DNA to replace material at damaged sites is identified by the *dinB* and *umuCD* genes that code for DNA polymerases IV and V in *E. coli*, the *rad30* gene coding for DNA polymerase η of *S. cerevisiae*, and the gene *XPV* that codes for the human homolog. They are sometimes called translesion DNA polymerases (for reveiw see 4523). A difference between the bacterial and eukaryotic enzymes is that the latter are not error-prone at thymine dimers: they accurately introduces an A-A pair opposite a T-T dimer (1167;

for review see 1164; 2940). When they replicate through other sites of damage, however, they are more prone to introduce errors.

An interesting feature of repair that has been best characterized in yeast is its connection with transcription. Transcriptionally active genes are preferentially repaired. The consequence is that the transcribed strand is preferentially repaired (removing the impediment to transcription). The cause appears to be a mechanistic connection between the repair apparatus and RNA polymerase. The RAD3 protein, which is a helicase required for the incision step, is a component of a transcription factor associated with RNA polymerase (see *Molecular Biology 5.21.12 A connection between transcription and repair*.

Mammalian cells show heterogeneity in the amount of DNA resynthesized at each lesion after damage. However, the patches are always relatively short, <10 bases.

An indication of the existence and importance of the mammalian repair systems is given by certain human hereditary disorders. The best investigated of these is xeroderma pigmentosum (XP), a recessive disease resulting in hypersensitivity to sunlight, in particular to ultraviolet. The deficiency results in skin disorders (and sometimes more severe defects).

The disease is caused by a deficiency in excision repair. Fibroblasts from XP patients cannot excise pyrimidine dimers and other bulky adducts. Mutations fall into 8 genes, called *XP-A* to *XP-G*. They have homologues in the *RAD* genes of yeast, showing that this pathway is a widely used in eukaryotes.

A protein complex that includes products of several of the *XP* genes is responsible for excision of thymine dimers. **Figure 15.53** shows its role in the repair pathway. The complex binds to DNA at a site of damage, perhaps by a mechanism involving the cooperative acion of several of its components 4519. Then the strands of DNA are unwound for ~20 bp around the damaged site. This action is undertaken by the helicase activity of the transcription factor $TF_{II}H$, itself a large complex, which includes the products of several *XP* genes, and which is involved with the repair of damaged DNA that is encountered by RNA polymerase during transcription. Then cleavages are made on either side of the lesion by endonucleases coded by *XP* genes. The single-stranded stretch including the damaged bases can then be replaced by synthesis of a replacement.

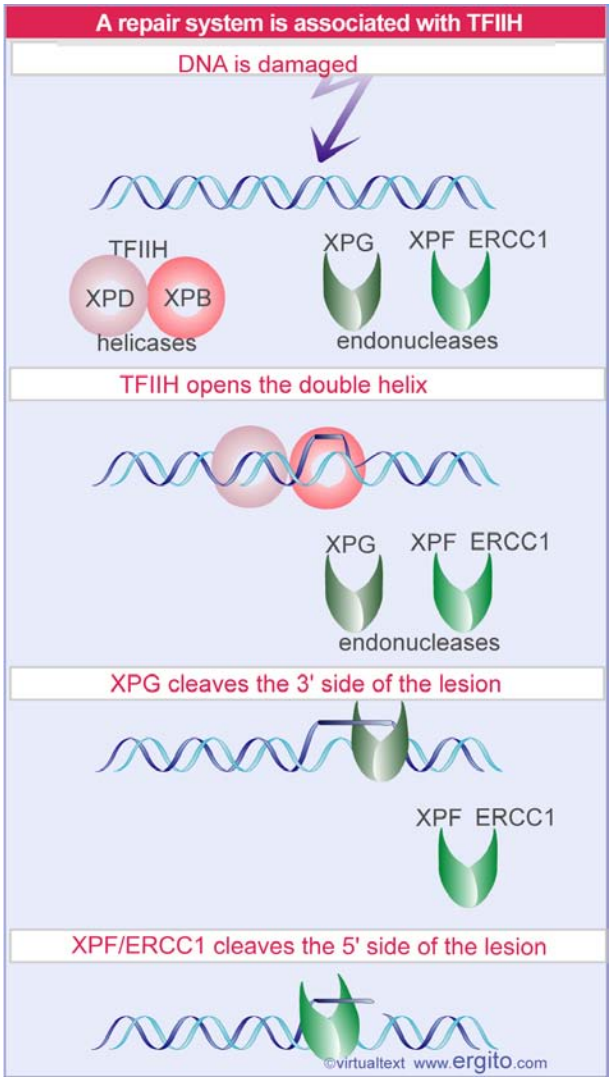

Figure 15.53 A helicase unwinds DNA at a damaged site, endonucleases cut on either side of the lesion, and new DNA is synthesized to replace the excised stretch.
This is a static version of an interactive figure; see http://www.ergito.com/main.jsp?bcs=MBIO.4.15.28 to view properly.

In cases where replication encounters a thymine dimer that has not been removed, replication requires the DNA polymerase η activity in order to proceed past the dimer. This is coded by the *XPV*. Skin cancers that occur in *XPV* mutants are presumably due to loss of the DNA polymerase.

Last updated on January 8, 2004

Reviews

2940. Prakash, S. and Prakash, L. (2002). *Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair.* Genes Dev. 16, 1872-1883.

References

1164. Friedberg, E. C., Feaver, W. J., and Gerlach, V. L. (2000). *The many faces of DNA polymerases: strategies for mutagenesis and for mutational avoidance*. Proc. Natl. Acad. Sci. USA 97, 5681-5683.

1167. Johnson, R. E., Prakash, S., and Prakash, L. (1999). *Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Pol eta.* Science 283, 1001-1004.

4190. Wolner, B., van Komen, S., Sung, P., and Peterson, C. L. (2003). *Recruitment of the recombinational repair machinery to a DNA double-strand break in yeast.* Mol. Cell 12, 221-232.

4519. Reardon, J. T. and Sancar, A. (2003). *Recognition and repair of the cyclobutane thymine dimer, a major cause of skin cancers, by the human excision nuclease.* Genes Dev. 17, 2539-2551.

4523. Rattray, A. J. and Strathern, J. N. (2003). *Error-prone DNA polymerases: when making a mistake is the only way to get ahead.* Annu. Rev. Genet. 37, 31-66.

RECOMBINATION AND REPAIR

4.15.29 A common system repairs double-strand breaks

Key Terms

Non-homologous end-joining (NHEJ) ligates blunt ends. It is common to many repair pathways and to certain recombination pathways (such as immunoglobulin recombination).

Key Concepts

- The NHEJ pathway can ligate blunt ends of duplex DNA.
- Mutations in the NHEJ pathway cause human diseases.

Double-strand breaks occur in cells in various circumstances. They initiate the process of homologous recombination and are an intermediate in the recombination of immunoglobulin genes (see *Molecular Biology 5.25.9 The RAG proteins catalyze breakage and reunion*). They also occur as the result of damage to DNA, for example, by irradiation. The major mechanism to repair these breaks is called **non-homologous end-joining** (**NHEJ**), and consists of ligating the blunt ends together.

The steps involved in NHEJ are summarized in **Figure 15.54**. The same enzyme complex undertakes the process in both NHEJ and immune recombination. The first stage is recognition of the broken ends by a heterodimer consisting of the proteins Ku70 and Ku80. They form a scaffold that holds the ends together and allows other enzymes to act on them. A key component is the DNA-dependent protein kinase (DNA-PKcs), which is activated by DNA to phosphorylate protein targets. One of these targets is the protein Artemis, which in its activated form has both exonuclease and endonuclease activities, and can both trim overhanging ends and cleave the hairpins generated by recombination of immunoglobulin genes (2397). The DNA polymerase activity that fills in any remaining single-stranded protrusions is not known. The actual joining of the double-stranded ends is undertaken by the DNA ligase IV, which functions in conjunction with the protein XRCC4. Mutations in any of these components may render eukaryotic cells more sensitive to radiation. Some of the genes for these proteins are mutated in patients who have diseases due to deficiencies in DNA repair.

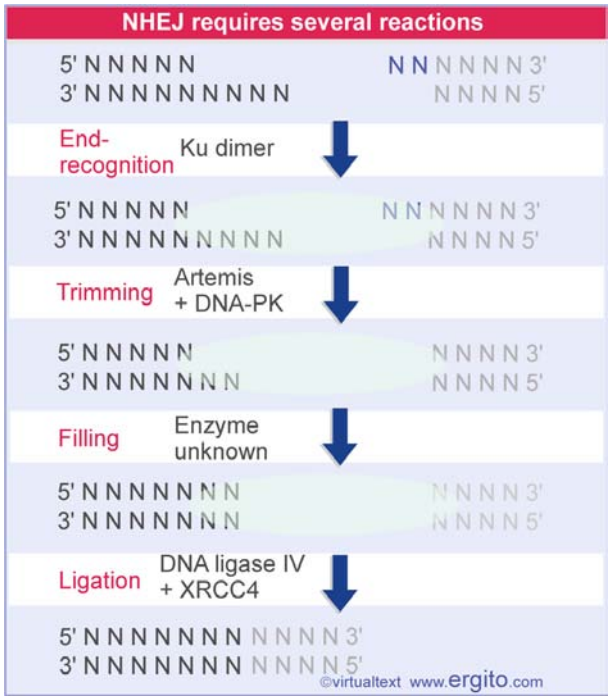

Figure 15.54 Nonhomologous end joining requires recognition of the broken ends, trimming of overhanging ends and/or filling, followed by ligation.

The Ku heterodimer is the sensor that detects DNA damage by binding to the broken ends. The crystal structure in **Figure 15.55** shows why it binds only to ends. The bulk of the protein extends for about two turns along one face of DNA (lower), but a narrow bridge between the subunits, located in the center of the structure, completely encircles DNA (2151). This means that the heterodimer needs to slip onto a free end.

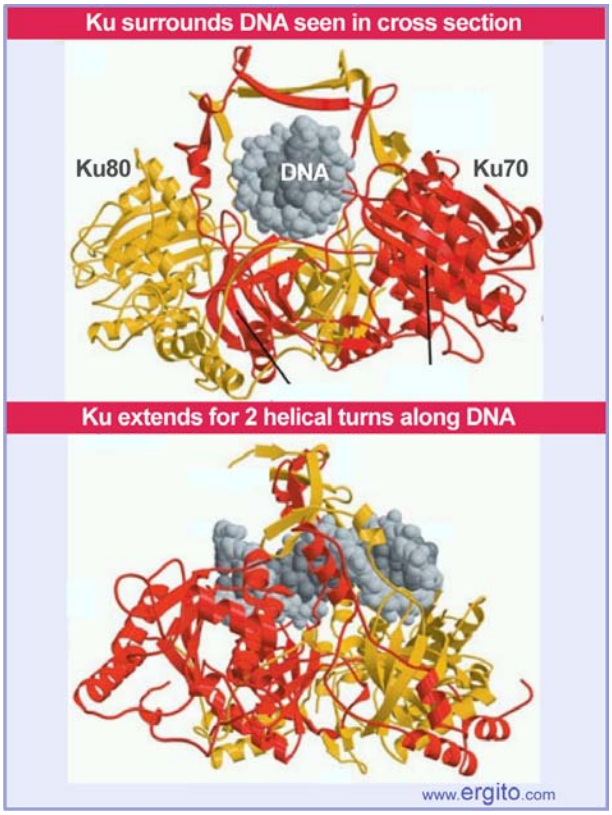

Figure 15.55 The Ku70-Ku80 heterodimer binds along two turns of the DNA double helix and surrounds the helix at the center of the binding site. Photograph kindly provided by Jonathan Goldberg (see 2151).

Ku can bring broken ends together by binding two DNA molecules (2152; 2153). The ability of Ku heterodimers to associate with one another suggests that the reaction might take place as illustrated in **Figure 15.56**. This would predict that the ligase would act by binding in the region between the bridges on the individual heterodimers. Presumably Ku must change its structure in order to be released from DNA.

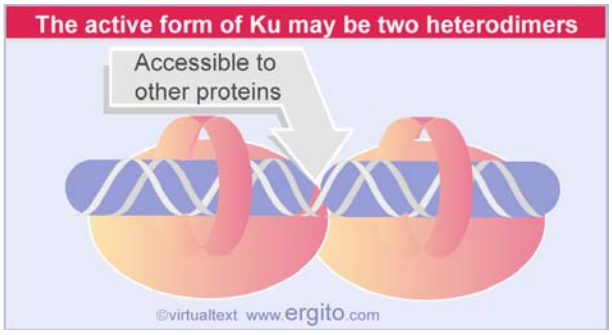

Figure 15.56 If two heterodimers of Ku bind to DNA, the distance between the two bridges that encircle DNA is ~12 bp.

Deficiency in DNA repair causes several human diseases. The common feature is

that an inability to repair double-strand breaks in DNA leads to chromosomal instability. The instability is revealed by chromosomal aberrations, which are associated with an increased rate of mutation, in turn leading to an increased susceptibility to cancer in patients with the disease. The basic cause can be mutation in pathways that control DNA repair or in the genes that code for enzymes of the repair complexes. The phenotypes can be very similar, as in the case of Ataxia telangiectasia (AT), which is caused by failure of a cell cycle checkpoint pathway, and Nijmegan breakage syndrome (NBS), which is caused by a mutation of a repair enzyme. One of the lessons that we learn from characterizing the repair pathways is that they are conserved in mammals, yeast, and bacteria.

The recessive human disorder of Bloom's syndrome is caused by mutations in a helicase gene (called *BLM*) that is homologous to *recQ* of *E. coli* (2927). The mutation results in an increased frequency of chromosomal breaks and sister chromatid exchanges. BLM associates with other repair proteins as part of a large complex. One of the proteins with which it interacts is hMLH1, a mismatch-repair protein that is the human homologue of bacterial *mutL*. The yeast homologues of these two proteins, Sgs1 and MLH1, also associate, identifying these genes as parts of a well-conserved repair pathway.

Nijmegan breakage syndrome results from mutations in a gene coding for a protein (variously called Nibrin, p95, or NBS1) that is a component of the Mre11/Rad50 repair complex (2925; 2926). Its involvement in repairing double-strand breaks is shown by the formation of foci containing the group of proteins when human cells are irradiated with agents that induce double-strand breaks. After irradiation, the kinase ATMP (coded by the *AT* gene) phosphorylates NBS1; this activates the complex, which localizes at sites of DNA damage. Subsequent steps involve triggering a checkpoint (a mechanism that prevents the cell cycle from proceeding until the damage is repaired; see *Molecular Biology 4.15.29 A common system repairs double-strand breaks*), and recruiting other proteins that are required to repair the damage (for review see 3431).

Last updated on 9-16-2002

Reviews

3431. D'Amours, D. and Jackson, S. P. (2002). *The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling.* Nat. Rev. Mol. Cell Biol. 3, 317-327.

References

2151. Walker, J. R., Corpina, R. A., and Goldberg, J. (2001). *Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair.* Nature 412, 607-614.

2152. Cary, R. B., Peterson, S. R., Wang, J., Bear, D. G., Bradbury, E. M., and Chen, D. J. (1997). *DNA looping by Ku and the DNA-dependent protein kinase.* Proc. Natl. Acad. Sci. USA 94, 4267-4272.

2153. Ramsden, D. A. and Gellert, M. (1998). *Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks.* EMBO J. 17, 609-614.

2397. Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M. R. (2002). *Hairpin Opening and Overhang Processing by an Artemis/DNA-Dependent Protein Kinase Complex in Nonhomologous End Joining and V(D)J Recombination.* Cell 108, 781-794.

2925. Varon, R. et al., (1998). *Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome.* Cell 93, 467-476.

2926. Carney, J. P., Maser, R. S., Olivares, H., Davis, E. M., Le Beau, M., Yates, J. R., Hays, L., Morgan, W. F., and Petrini, J. H. (1998). *The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response.* Cell 93, 477-486.

2927. Ellis, N. A., Groden, J., Ye, T. Z., Straughen, J., Lennon, D. J., Ciocci, S., Proytcheva, M., and German, J. (1995). *The Bloom's syndrome gene product is homologous to RecQ helicases.* Cell 83, 655-666.

RECOMBINATION AND REPAIR

4.15.30 Summary

Recombination involves the physical exchange of parts between corresponding DNA molecules. This results in a duplex DNA in which two regions of opposite parental origins are connected by a stretch of hybrid (heteroduplex) DNA in which one strand is derived from each parent. Correction events may occur at sites that are mismatched within the hybrid DNA. Hybrid DNA can also be formed without recombination occurring between markers on either side. Gene conversion occurs when an extensive region of hybrid DNA forms during normal recombination (or between nonallelic genes in an aberrant event) and is corrected to the sequence of only one parental strand; then one gene takes on the sequence of the other.

Recombination is initiated by a double-strand break in DNA. The break is enlarged to a gap with a single-stranded end; then the free single-stranded end forms a heteroduplex with the allelic sequence. The DNA in which the break occurs actually incorporates the sequence of the chromosome that it invades, so the initiating DNA is called the recipient. Hot spots for recombination are sites where double strand breaks are initiated. A gradient of gene conversion is determined by the likelihood that a sequence near the free end will be converted to a single strand; this decreases with distance from the break.

Recombination is initiated in yeast by Spo11, a topoisomerase-like enzyme that becomes linked to the free 5′ ends of DNA. The DSB is then processed by generating single-stranded DNA that can anneal with its complement in the other chromosome. Yeast mutations that block synaptonemal complex formation show that recombination is required for its formation. Formation of the synaptonemal complex may be initiated by double-strand breaks, and it may persist until recombination is completed. Mutations in components of the synaptonemal complex block its formation but do not prevent chromosome pairing, so homologue recognition is independent of recombination and synaptonemal complex formation.

The full set of reactions required for recombination can be undertaken by the Rec and Ruv proteins of *E. coli*. A single stranded region with a free end is generated by the RecBCD nuclease. The enzyme binds to DNA on one side of a *chi* sequence, and then moves to the *chi* sequence, unwinding DNA as it progresses. A single-strand break is made at the *chi* sequence. *chi* sequences provide hotspots for recombination. The single-strand provides a substrate for RecA, which has the ability to synapse homologous DNA molecules by sponsoring a reaction in which a single strand from one molecule invades a duplex of the other molecule. Heteroduplex DNA is formed by displacing one of the original strands of the duplex. These actions create a recombination junction, which is resolved by the Ruv proteins. RuvA and RuvB act at a heteroduplex, and RuvC cleaves Holliday junctions.

Recombination, like replication and (probably) transcription, requires topological manipulation of DNA. Topoisomerases may relax (or introduce) supercoils in DNA, and are required to disentangle DNA molecules that have become catenated by recombination or by replication. Type I topoisomerases introduce a break in one strand of a DNA duplex; type II topoisomerases make double-stranded breaks. The

enzyme becomes linked to the DNA by a bond from tyrosine to either 5′ phosphate (type A enzymes) or 3′ phosphate (type B enzymes).

The enzymes involved in site-specific recombination have actions related to those of topoisomerases. Among this general class of recombinases, those concerned with phage integration form the subclass of integrases. The Cre-*lox* system uses two molecules of Cre to bind to each *lox* site, so that the recombining complex is a tetramer. This is one of the standard systems for inserting DNA into a foreign genome. Phage lambda integration requires the phage Int protein and host IHF protein and involves a precise breakage and reunion in the absence of any synthesis of DNA. The reaction involves wrapping of the *attP* sequence of phage DNA into the nucleoprotein structure of the intasome, which contains several copies of Int and IHF; then the host *attB* sequence is bound, and recombination occurs. Reaction in the reverse direction requires the phage protein Xis. Some integrases function by *cis*-cleavage, where the tyrosine that reacts with DNA in a half site is provided by the enzyme subunit bound to that half site; others function by *trans*-cleavage, where a different protein subunit provides the tyrosine.

Bacteria contain systems that maintain the integrity of their DNA sequences in the face of damage or errors of replication and that distinguish the DNA from sequences of a foreign source.

Repair systems can recognize mispaired, altered, or missing bases in DNA, or other structural distortions of the double helix. Excision repair systems cleave DNA near a site of damage, remove one strand, and synthesize a new sequence to replace the excised material. The Uvr system provides the main excision-repair pathway in *E. coli*. The *dam* system is involved in correcting mismatches generated by incorporation of incorrect bases during replication and functions by preferentially removing the base on the strand of DNA that is not methylated at the *dam* target sequence. Eukaryotic homologues of the *E. coli* MutSL system are involved in repairing mismatches that result from replication slippage; mutations in this pathway are common in certain types of cancer.

Recombination-repair systems retrieve information from a DNA duplex and use it to repair a sequence that has been damaged on both strands. The RecBC and RecF pathways both act prior to RecA, whose strand-transfer function is involved in all bacterial recombination. A major use of recombination-repair may be to recover from the situation created when a replication fork stalls.

The other capacity of recA is the ability to induce the SOS response. RecA is activated by damaged DNA in an unknown manner. It triggers cleavage of the LexA repressor protein, thus releasing repression of many loci, and inducing synthesis of the enzymes of both excision repair and recombination-repair pathways. Genes under LexA control possess an operator SOS box. RecA also directly activates some repair activities. Cleavage of repressors of lysogenic phages may induce the phages to enter the lytic cycle.

Repair systems can be connected with transcription in both prokaryotes and eukaryotes. Human diseases are caused by mutations in genes coding for repair activities that are associated with the transcription factor TFIIH. They have homologues in the *RAD* genes of yeast, suggesting that this repair system is

widespread.

Nonhomologous end joining (NHEJ) is a general reaction for repairing broken ends in (eukaryotic) DNA. The Ku heterodimer brings the broken ends together so they can be ligated. Several human diseases are caused by mutations in enzymes of this pathway.