Type of Recombinations

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.
Specialized/site-specific recombination occurs only between specific sites. The results depend on the locations of the two recombining sites.
An intramolecular recombination between two specific sites on a circular DNA releases two smaller circular DNAs.
Visible changes in chromosomes through the five stages of meiotic prophase that are involved in exchanging material between duplexes of DNA.

**Homologous Recombination Between synapsed (paired) chromosomes – during Meiosis**
- **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.
There is no detailed information about the molecular events involved in recombination in higher eukaryote.

However, recently the isolation of mutants in yeast has made it possible to correlate some of the molecular steps with approximate stages of meiosis.

Most of the current knowledge is concerned with recombination in bacteria.
Possible outcomes of recombination
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.

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**.
Branch migration can occur in either direction and 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*.

Branch migration could allow the point of crossover in the recombination intermediate to move in either direction.
How is joint molecule resolved? Resolution of joint molecule requires a further pair of nicks.

If the nicks are made in the pair of strands that were not originally nicked, all four of the original strands have been nicked. This releases splice recombinant DNA molecules.

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.
What is the minimum length of the region required to establish the connection between the recombining duplexes?

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.
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.

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.
The chromosomes enter meiosis and pair to form the **synaptonemal complex**. It has been assumed that this represents some stage involved with recombination. A more recent view is that the synaptonemal complex is a consequence rather than a cause of recombination.

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**.
The distinctive structural features of synaptonemal complex are due to two groups of proteins:

1. **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.
2. The lateral elements are connected by transverse filaments that are equivalent to the central element. These are formed from Zip proteins.
In *rad50* mutants (yeast), 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 that generates the double-strand breaks.

The 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.
In yeast double-strand breaks appear and then disappear over a 60 minute period. The first joint molecules 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.

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 \textit{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.
Bacterial DNA Recombination Systems

- Bacteria do not usually exchange large amounts of duplex DNA.

- 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 \textit{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 \textit{chi} sequence occurs naturally in \textit{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, 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. 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*.

*recBCD* 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.
Rec mutations of *E. coli* cannot undertake general recombination.

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. 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, 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. 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.
RecA catalyzes strand exchange between duplex and single-stranded molecules
RecA has two quite different types of activity: it can stimulate protease activity in the SOS response 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.
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. 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.
Holliday Structure is Resolved by Ruv Proteins

The reaction between a partially duplex molecule and an entirely duplex molecule leads to the exchange of strands (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.
RuvAB catalyzes branch migration

RuvB hexamer binds as ring around DNA

RuvA tetramer contacts all 4 strands

Branch migration
The proteins involved in stabilizing and resolving Holliday junctions have been identified as the products of the *Ruv* genes in *E. coli*.

- 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.
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).

All of this suggests that recombination uses a "resolvasome" complex that includes enzymes catalyzing branch
Summary of 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.
Supercoiled circular molecule that forms a twisted and therefore more condensed shape.
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.
A DNA topoisomerase is an enzyme that changes the number of times the two strands in a closed DNA molecule cross each other. It does this by cutting the DNA, passing DNA through the break, and resealing the DNA.

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.
**Single-strand passage** is a reaction catalyzed by type I topoisomerase in which one section of single-stranded DNA is passed through another strand.
Type II topoisomerases relax both negative and positive supercoils. The reaction requires ATP, with one ATP hydrolyzed for each catalytic event. 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.
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.

Gyrase protects ~140 bp of DNA from digestion by micrococcal nuclease. 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.
Site specific recombination
- **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.

- **att** sites are the loci on a lambda phage and the bacterial chromosome at which recombination integrates the phage into, or excises it from, the bacterial chromosome.

- 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** (O) 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.
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.
The function of att$P$ requires a stretch of 240 bp, but the function of att$B$ 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 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. 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.
Lambda recombination complex is an intasome

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 *λint*, 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*.

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. The two domains probably bind DNA simultaneously, thus bringing the arms of *attP* close to the core.

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**.

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.
Repair systems correct damage to DNA

- 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.
Single base changes may occur due to deamination of cytosine to uracil (spontaneously or by chemical mutagen) creates a mismatched U•G pair. So only limited time is available to repair the damage before it is fixed by replication.
Replication errors introduce mismatched base pairs

Nature of mutation

Cytosine

\[
\begin{array}{c}
\text{H} \\
\text{N} \\
\text{H} \\
\end{array} \\
\begin{array}{c}
\text{N} \\
\text{O} \\
\end{array}
\]

Replication errors

\[
\begin{array}{c}
\text{H} \\
\text{N} \\
\text{H} \\
\end{array} \\
\begin{array}{c}
\text{N} \\
\text{O} \\
\end{array}
\]

Adenine

Consequences

Purine pair distorts duplex

Corrected by removing A or G in newly synthesized strand

©virtualtext www.ergito.com
Structural distortions:

Figure 15.34 shows the example of ultraviolet irradiation, which introduces covalent bonds between two adjacent thymine bases, giving an intrastrand pyrimidine dimer.
Methylation can distort the structure of DNA

**Nature of mutation**
- Guanine
- Methyl-guanine

**Consequences**
- Methyl group distorts double helix
- Corrected by dealkylation

©virtualtext  www.ergito.com
Depurination requires base replacement

Nature of mutation

Adenine

Depurination

Consequences

Purine is missing

Corrected by insertion

©virtualtext  www.ergito.com
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.

These genes may be divided into several general types, as summarized in Figure 15.37:
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.. 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.
Nucleotide Excision Repair

Excision-repair replaces damaged DNA

DNA is damaged

Damage is removed

Replacement DNA is synthesized
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 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.

**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.
In E. coli, known repair 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.
Excision Repair System in *E. coli*

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 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).
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).

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 (A connection between transcription and repair).
Single Base Removal

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. A glycosylase cleaves the bond between the damaged or mismatched base and the deoxyribose.
Some glycosylases are also lyases that can take the reaction a stage further by using an amino (NH2) 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.
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. 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.
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.

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

By contrast with this mechanism, 1-methyl-adénylne 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 CH2OH group, and then the release of the HCHO moiety [formaldehyde] restores the structure of adénine. 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. 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.

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.
Damaged DNA that has not been repaired causes DNA polymerase III to stall during replication.

DNA polymerase V (coded by \textit{umuCD}), or DNA polymerase IV (coded by \textit{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 *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 8-oxo-G 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.
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.
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.
*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.

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.
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.
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. 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. They are repaired by homologues of the MutSL system as shown in Figure 15.47.
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. An example is illustrated in Figure 15.48. Restarting stalled replication forks could be the major role of the recombination-repair systems.
Recombination is an important mechanism to recover from replication errors.

**Figure 15.49** shows one possible outcome when a replication fork stalls. 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. DNA polymerase II is required for the replication restart, and is later replaced by DNA polymerase III.
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.
RecA triggers the SOS system

Activation of RecA causes proteolytic cleavage of the product of the \textit{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.
Eukaryotic cells have conserved repair systems

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 action of several of its components. 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 TFIIR, 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.