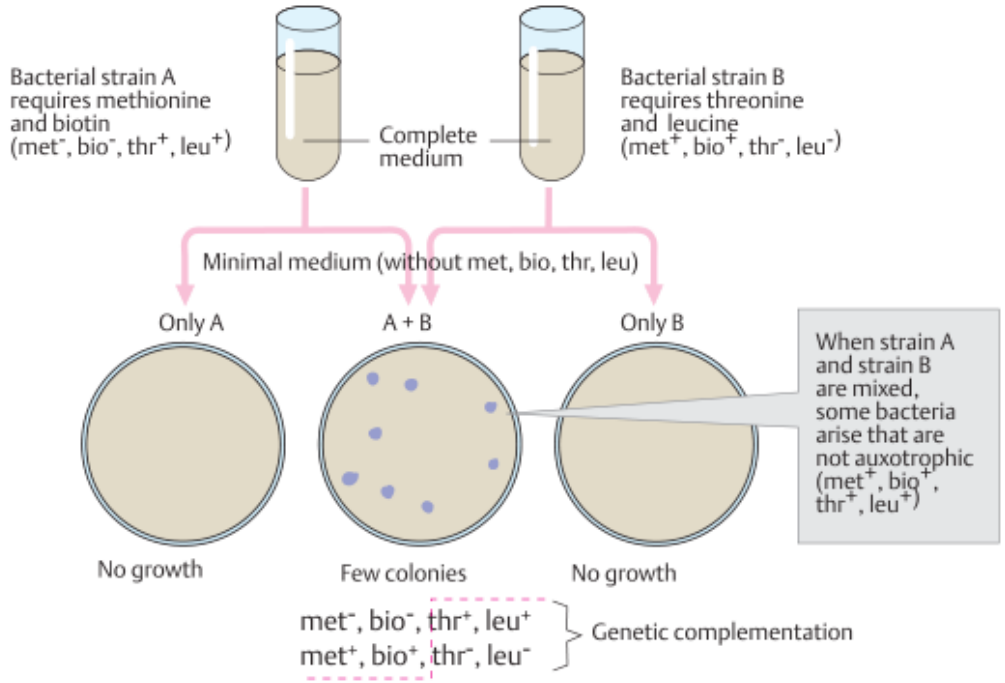
Recombination in Bacteria

In 1946, J. Lederberg and E. L. Tatum first demonstrated that genetic information can be exchanged between different mutant bacterial strains. This corresponds to a type of sexuality and leads to genetic recombination.

A. Genetic recombination in bacteria

In their classic experiment, Lederberg and Tatum used two different auxotrophic bacterial strains. One (A) was auxotrophic for methionine (Met –) and biotin (Bio –). This strain required methionine and biotin, but not threonine and leucine (Thr + , Leu +), to be added to the medium. The opposite was true for bacterial strain B, auxotrophic for threonine and leucine (Thr – , Leu –), but prototrophic for methionine and biotin (Met + , Bio +). When the cultures were mixed together without the addition of any of these four amino acids and then plated on an agar plate with minimal medium, a few single colonies (Met + , Bio + , Thr + , Leu +) unexpectedly appeared. Although this occurred rarely (about 1 in 10 7 platedcells), a few colonies with altered genetic properties usually appeared owing to the large number of plated bacteria. The interpretation: genetic recombination between strain A and strain B. The genetic properties of the parent cells complemented each other (genetic complementation). (Figure adapted from Stent & Calendar, 1978).

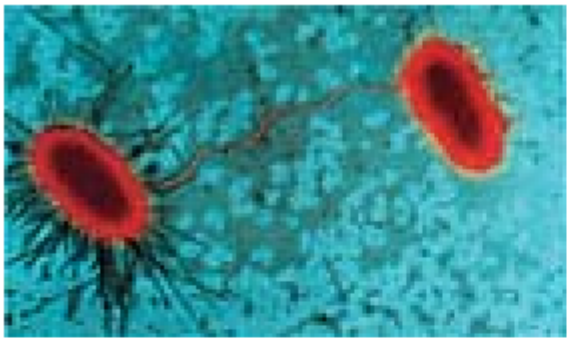


Genetic recombination in bacteria

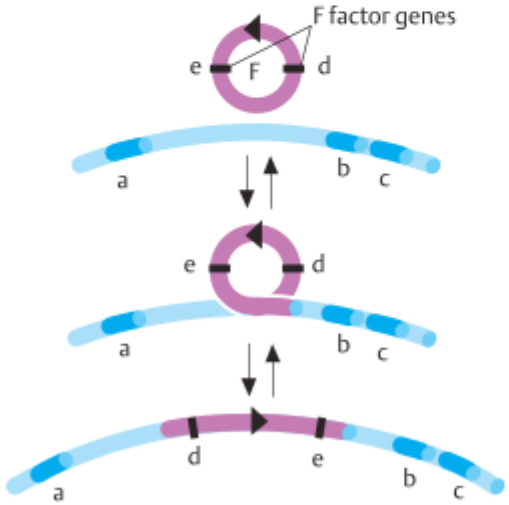
B. Conjugation in bacteria

Later, the genetic exchange between bacteria (conjugation) was demonstrated by light microscopy. Conjugation occurs with bacteria possessing a gene that enables frequent recombination. Bacterial DNA transfer occurs in one direction only. “Male” chromosomal material is introduced into a “female” cell. The so-called male and female cells of E.coli differ in the presence of a fertility factor (F). When F + and F – cells are mixed together, conjugal pairs can form with attachment of a male (F +) sex pilus to the surface of an F – cell. (Photograph from Science 257 :1037, 1992). C. Integration of the F factor into an Hfr – chromosome The F factor can be integrated into the bacterial chromosome by means of specific crossingover. After the factor is integrated, the original bacterial chromosome with the sections a, b, and c contains additional genes, the F factor genes(e,d). Such a chromosome is called an Hfr chromosome (Hfr, high frequency of recombination) owing to its high rate of recombination with genes of other cells as a result of conjugation. D. Transfer of F DNA from an F + to an F – cell

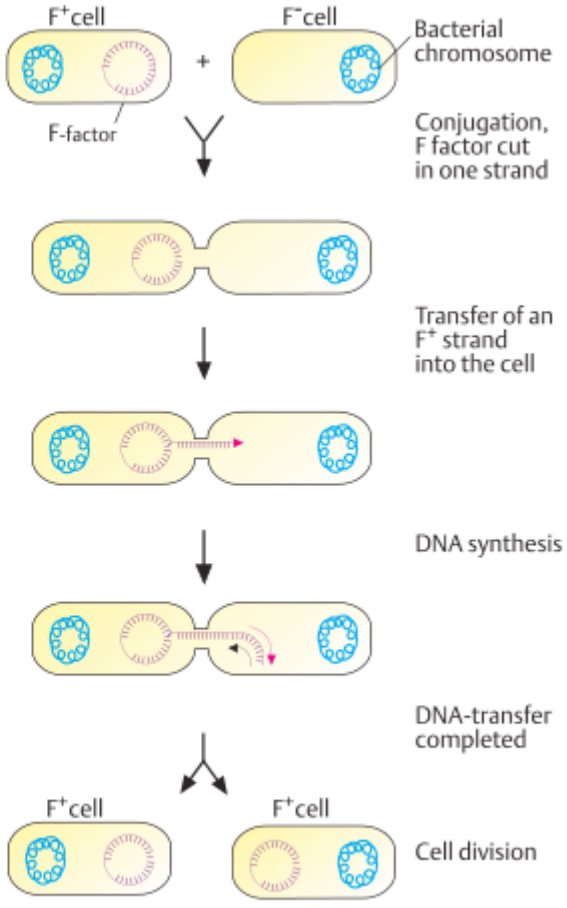
Bacteria may contain the F factor(fertility) as an additional small chromosome, i.e., a small ring shaped DNA molecule (F plasmid) of about 94000 base pairs (not shown to scale). This corresponds to about 1/40th of the total genetic information of a bacterial chromosome. It occurs once per cell and can be transferred to other bacterial cells. About a third of the F + DNA consists of transfer genes, including genes for the formation of sex pili. The transfer of the F factor begins after a strand of the DNA double helix is opened. One strand is transferred to the acceptor cell. There it is replicated, so that it becomes double-stranded. The DNA strand remaining in the donor cell is likewise restored to a double strand by replication. Thus, DNA synthesis occurs in both the donor and the acceptor cell. When all is concluded, the acceptor cell is also an F + cell.(Figures in C and D adapted from Watson et al., 1987).



B. Conjugation in bacteria



C. Integration of the F factor into an Hfr⁻ chromosome



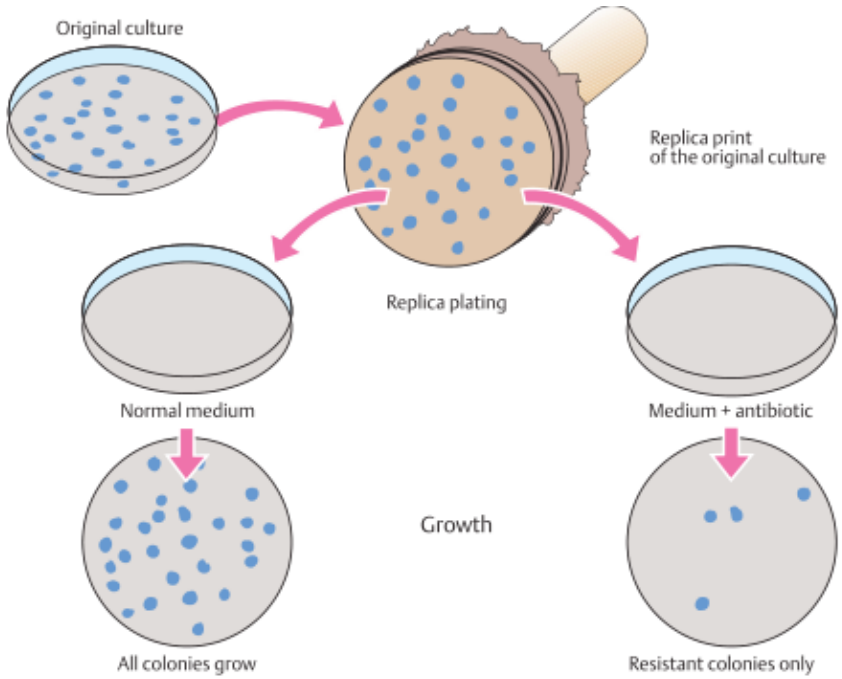
D. Transfer of F DNA from an F⁺ into an F⁻ cell

Isolation of Mutant Bacteria

Important advances in genetics were made in the early 1950s through studies of bacteria. As prokaryotic organisms, bacteria have certain advantages over eukaryotic organisms because they are haploid and have an extremely short generation time. Mutant bacteria can be identified easily. The growth of some mutant bacteria depends on whether a certain substance is present in the medium (auxotrophism). Bacterial cultures are well suited for determining mutational events since an almost unlimited number of cells can be tested in a short time. Without great difficulty, it is possible to detect one mutant in 10e7 colonies. Efficiency to this degree is not possible in the genetic analysis of eukaryotic organisms.

A. Replica plating to recognize mutants

In 1952, Joshua and Esther Lederberg developed replica plating of bacterial cultures. With this method, individual colonies on an agar plate can be taken up with a stamp covered with velvet and placed onto other culture dishes with media of different compositions. Some mutant bacteria differ from non mutants in their ability to grow. Here several colonies are shown in the Petri dish of the initial culture. Each of these colonies originated from a single cell. By means of replica plating, the colonies are transferred to two new cultures. One culture (right) contains an antibiotic in the culture medium; the other (left) does not. All colonies grow in normal medium, but only those colonies that are antibiotic resistant owing to a mutation grow in the antibiotic-containing medium. In this manner, mutant colonies can be readily identified.

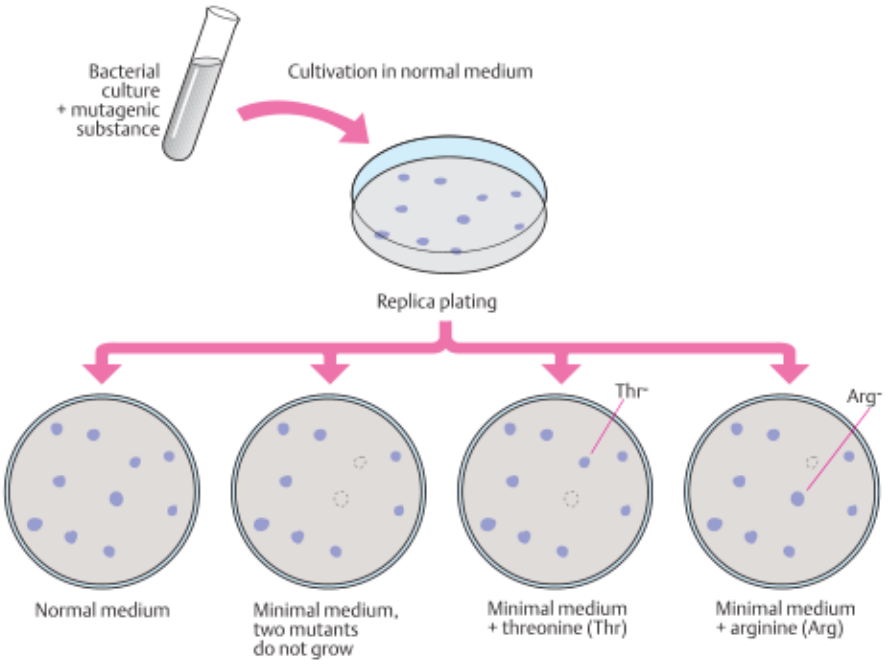


Replica plating to recognize mutants

B. Mutant bacteria identified through an auxotrophic medium

Here it is shown how different mutants can be distinguished, e.g., after exposure to a mutagenic substance. After a colony has been treated with a mutagenic substance, it is first cultivated in normal nutrient medium. Mutants can then be identified by replica plating. The culture with the normal medium serves as the control. In one culture with minimal medium, from which a number of substances are absent, two colonies do not grow (auxotrophic mutants). Initially, it is known for which of the substances the colonies are auxotrophic. If a different amino acid is added to each of two cultures with minimal medium, e.g., threonine (Thr) to one and arginine (Arg) to the other, it can be observed that one of the mutant colonies grows in the threonine-containing minimal medium, but the other does not. The former colony is dependent on the presence of threonine (Thr –), i.e., it is an auxotroph for threonine. The other culture with minimal medium had arginine added. Only here can the other of the two mutant colonies, an auxotroph for arginine (Arg –), grow. After the

mutant colonies requiring specific conditions for growth have been identified, they can be further characterized. This procedure is relatively simple and makes rapid identification of mutants possible. Many mutant bacteria have been defined by auxotrophism. The wild-type cells that do not have special additional growth requirements are called prototrophs (Figures adapted from Stent & Calendar, 1978).



Mutant bacteria identified through an auxotrophic medium

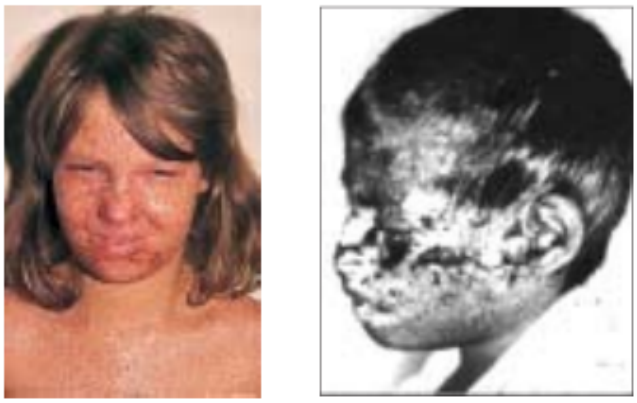
Xeroderma Pigmentosum

Xeroderma pigmentosum (XP) is a heterogeneous group of genetically determined skin disorders due to unusual sensitivity to ultraviolet light. They are manifested by dryness and pigmentation of the exposed regions of skin (xeroderma pigmentosum=“dry, pigmented skin”). The exposed areas of skin also show a tendency to develop tumors. The causes are different genetic defects of DNA repair. Repair

involves mechanisms similar to those involved in transcription and replication. The necessary enzymes are encoded by at least a dozen genes, which are highly conserved in bacteria, yeast, and mammals.

A. Clinical phenotype

The skin changes are limited to UV-exposed areas (2). Unexposed areas show no changes. Thus it is important to protect patients from UV light. An especially important feature is the tendency for multiple skin tumors to develop in the exposed areas (3). These may even occur in childhood or early adolescence. The types of tumors are the same as those occurring in healthy individuals after prolonged UV exposure.

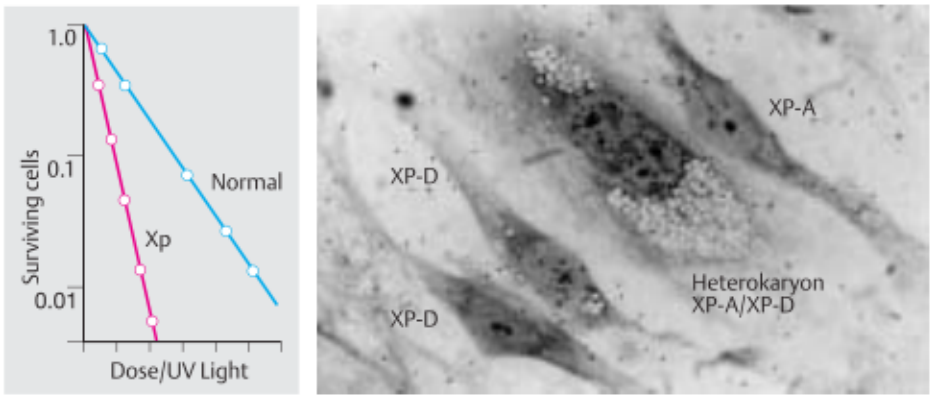


Clinical phenotype

B. Cellular phenotype

The UV sensitivity of cells can be demonstrated in vitro. When cultured fibroblasts from the skin of patients are exposed to UV light, the cells show a distinct dose-dependent decrease in survival rate compared with normal cells (1). Different degrees of UV sensitivity can be demonstrated. The short segment of new DNA normally formed during excision repair can be demonstrated by culturing cells in the presence of [3 H]thymidine and exposing them to UV light. The DNA synthesis induced for DNA repair can be made visible in autoradiographs.

Since [3 H]thymidine is incorporated during DNA repair, these bases are visible as small dots caused by the isotope on the film (2). In contrast, xeroderma (XP) cells show markedly decreased or almost absent repair synthesis. (Photograph of Bootsma & Hoeijmakers, 1999).

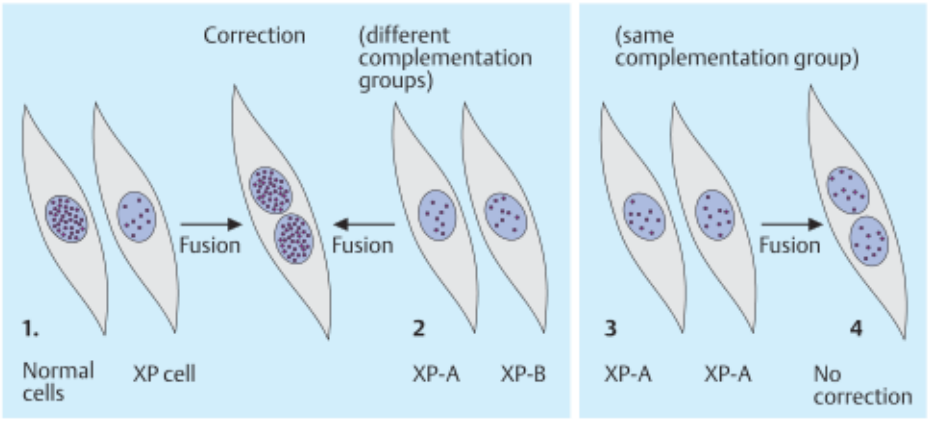


1. UV sensitivity **2.** Complementation following fusion of XP-A and XP-D cells

Cellular phenotype

C. Genetic complementation in cell hybrids

If skin cells (fibroblasts) from normal persons and from patients (XP) are fused (cell hybrids) in culture and exposed to UV light, the cellular XP phenotype will be corrected (1). Normal DNA repair occurs. Also, hybrid cells from two different forms of XP show normal DNA synthesis (2) because cells with different repair defects correct each other (genetic complementation). However, if the mutant cells have the same defect (3), they are not be able to correct each other (4) because they belong to the same complementation group. At present about ten complementation groups are known in xeroderma pigmentosum. They differ clinically in terms of severity and central nervous system involvement. Each complementation group is based on a mutation at a different gene locus. Several of these genes have been cloned and show homology with repair genes of other organisms, including yeast and bacteria.



Genetic complementation in cell hybrids

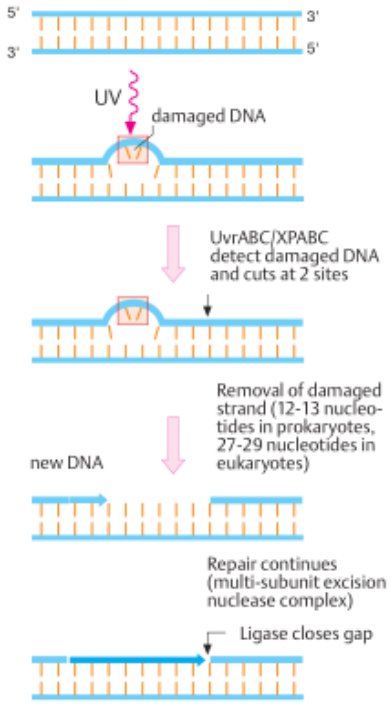
DNA Repair

Life would not be possible without the ability to repair damaged DNA. Since replication errors, including mismatch, and harmful exogenous factors are everyday problems for a living organism, a broad repertoire of repair genes has evolved in prokaryotes and eukaryotes. The following types of DNA repair can be distinguished by their basic mechanisms: (1) excision repair to remove a damaged DNA site, such as a strand with a thymine dimer; (2) mismatch repair to correct errors of replication by excising a stretch of single-stranded DNA containing the wrong base; (3) repair of UV-damaged DNA during replication; and (4) transcription coupled repair in active genes.

A. Excision repair

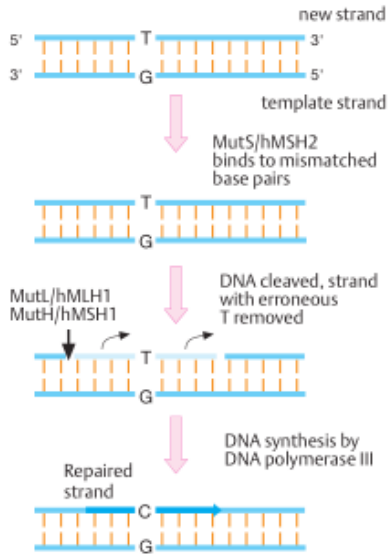
The damaged strand of DNA is distorted and can be recognized by a set of three proteins, the UvrA, UvrB, and UvrC endonucleases in prokaryotes and XPA, XPB, and XPC in human cells. This DNA strand is cleaved on both sides of the damage by an exonuclease protein complex, and a stretch of about 12

or 13 nucleotides in prokaryotes and 27 to 29 nucleotides in eukaryotes is removed. DNA repair synthesis restores the missing stretch and a DNA ligase closes the gap.



Excision repair

B. Mismatch repair

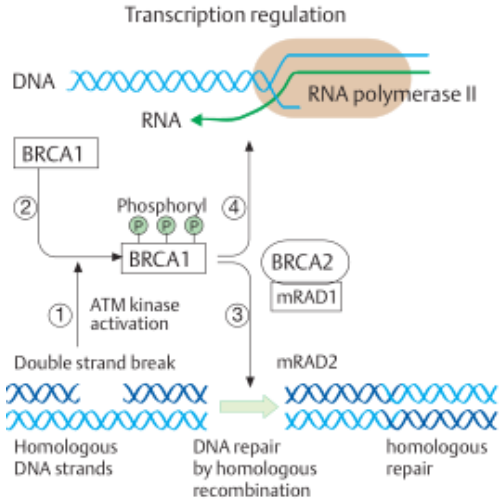


Mismatch repair

Mismatch repair corrects errors of replication. However, the newly synthesized DNA strand containing the wrong base must be

distinguished from the parent strand, and the site of a mismatch identified. The former is based on a difference in methylation in prokaryotes. The daughter strand is undermethylated at this stage. E.coli has three mismatch repair systems: long patch, short patch, and very short patch. The long patch system can replace 1kb DNA and more. It requires three repair proteins, MutH, MutL, and MutS, which have the human homologues hMSH1, hMLH1, and hMSH2. Mutations in their respective genes lead to cancer due to defective mismatch repair. C. Replication repair of UV-damaged DNA DNA damage interferes with replication, especially in the leading strand. Large stretches remain unreplicated beyond the damaged site (in the 3? direction of the new strand) unless swiftly repaired. The lagging strand is not affected as much because Okazaki fragments (about 100 nucleotides in length) of newly synthesized DNA are also formed beyond the damaged site. This leads to an asymmetric replication fork and single-stranded regions of the leading strand. Aside from repair by recombination, the damaged site can be bypassed.

D. Double-strand repair by homologous recombination



Double-strand repair by homologous recombination

Double-strand damage is a common consequence of γ radiation. An important human pathway for mediating repair requires three proteins, encoded by the genes ATM, BRCA1, and BRCA2. Their names are derived from important diseases that result from mutations in these genes: ataxia telangiectasia and hereditary predisposition to breast cancer (BRCA1 and BRCA2). ATM, a member of a protein kinase family, is activated in response to DNA damage (1). Its active form phosphorylates BRCA1at specific sites (2). Phosphorylated BRCA1 induces homologous recombination in cooperation with BRCA2 and mRAD5, the mammalian homologue of E. coli RecA repair protein (3). This is required for efficient DNA double-break repair. Phosphorylated BRCA1 may also be involved in transcription and transcription-coupled DNA repair (4). (Figure redrawn from Ventikaraman, 1999).

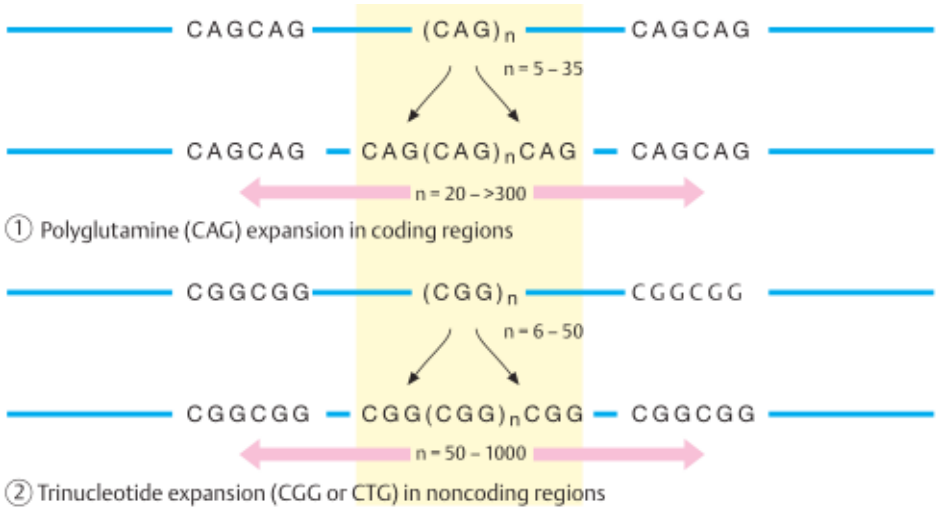
Trinucleotide Repeat Expansion

The human genome contains tandem repeats of trinucleotides. Normally they occur in groups of 5–35 repeats. When their number exceeds a certain threshold and they occur in a gene or close to it, they cause diseases. Once the normal, variable length has expanded, the increased number of repeats tends to increase even further when passed through the germline or during mitosis. Thus, trinucleotide expansions form a class of unstable mutations, to date observed in humans only.

A. Different types of trinucleotide

repeats and their expansions

Trinucleotide repeats can be distinguished according to their localization with respect to a gene. Expansions are greater outside genes and more moderate within coding regions. In several severe neurological diseases, abnormally expanded CAG repeats are part of the gene. CAG repeats encode a series of glutamines (polyglutamine tracts). Within a normal number of repeats, which varies according to the gene involved, the gene functions normally. However, an expanded number of repeats leads to an abnormal gene product with altered function. Trinucleotide repeats also occur in noncoding regions of a gene. Fairly common types are CGG and GCC repeats. The increase in the number of these repeats can be drastic, up to 1000 or more repeats. The first stages of expansion usually do not lead to clinical signs of a disease, but they do predispose to increased expansion of the repeat in the off spring of a carrier (premutation).

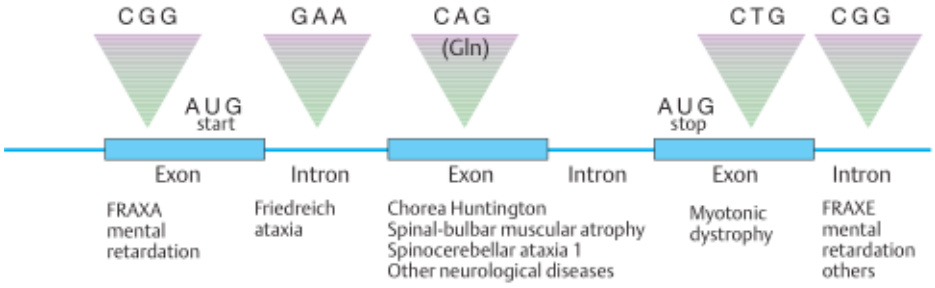


Different types of trinucleotide repeat expression

B. Unstable trinucleotide repeats

in different diseases

Disorders due to expansion of trinucleotide repeats can be distinguished according to the type of trinucleotide repeat,i.e., the sequence of the three nucleotides, their location with respect to the gene involved, and their clinical features. All involve the central or the peripheral nervous system. Type I trinucleotide diseases are characterized by CAG trinucleotide expansion within the coding region of different genes. The triplet CAG codes for glutamine. About 20 CAG repeats occur normally in these genes, so that about 20 glutamines occur in the gene product. In the disease state the number of glutamines is greatly increased in the protein. Hence, they are collectively referred to as polyglutamine disorders. Type II trinucleotide diseases are characterized by expansion of CTG, GAA, GCC, or CGG trinucleotides within a noncoding region of the gene involved, either at the 5′ end (GCC in fragile X syndrome type A, FRAXA), at the 3′ end (CGG in FRAXE; CTG in myotonic dystrophy), or in an intron (GAA in Friedreich ataxia).

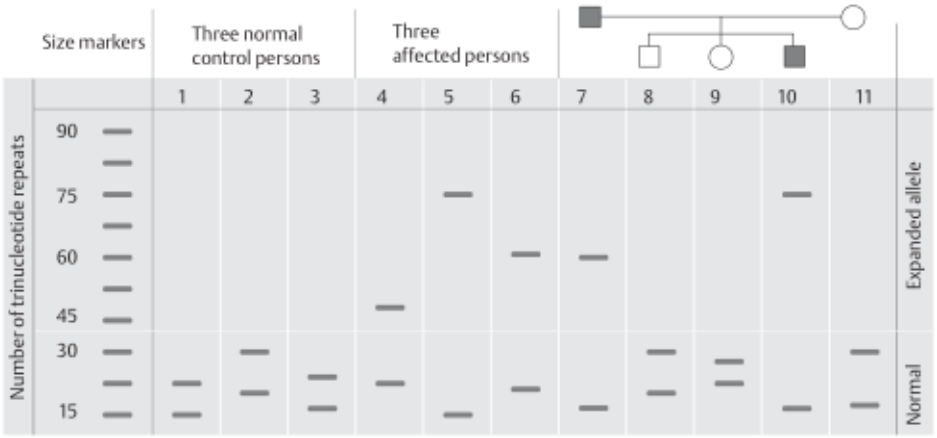


Unstable trinucleotide repeats in different diseases

C. Principle of laboratory diagnosis of unstable trinucleotide

repeats

The laboratory diagnosis compares the sizes of the trinucleotide repeats in the two alleles of the gene examined. One can distinguish very large expansions of repeats outside coding sequences (50 to more than 1000 repeats) and moderate expansion within coding sequences (20 to 100–200). The figure shows 11 lanes, each representing one person: normal controls in lanes 1–3; confirmed patients in lanes 4–6; and a family with an affected father (lane 7), an affected son (lane 10), the unaffected mother (lane 11), and two unaffected children: a son (lane 8) and a daughter (lane 9). Size markers are shown at the left. Each lane represents a polyacrylamide gel and the (CAG) n repeat of the Huntington locus amplified by polymerase chain reaction shown as a band of defined size. Each person shows the two alleles. In the affected persons the band representing one allele lies above the threshold in the expanded region (in practice the bands are somewhat blurred because the exact repeat size varies in DNA from different cells).



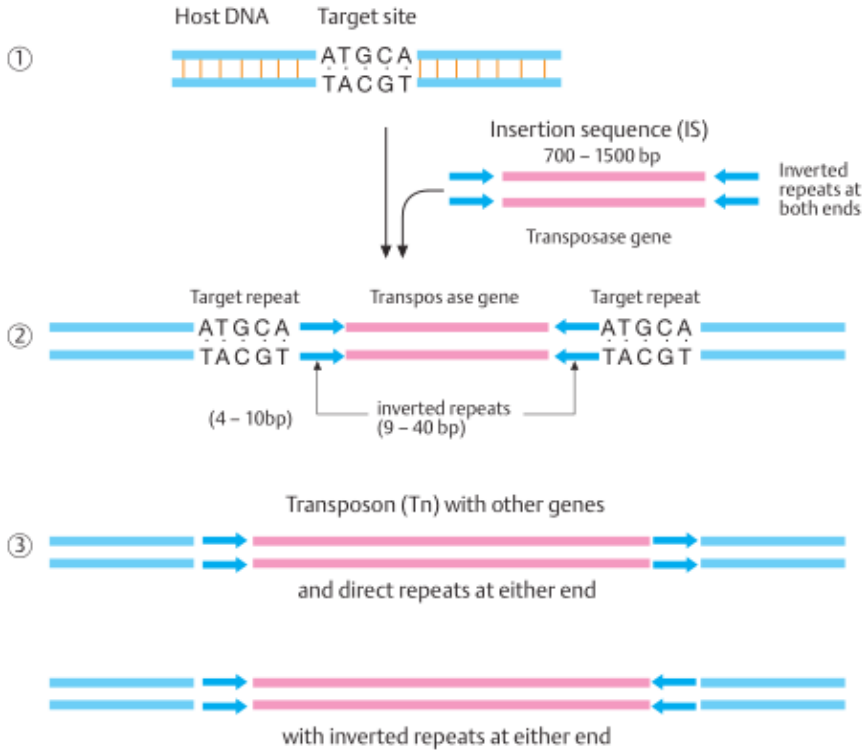
Principle of laboratory diagnosis of unstable trinucleotide repeats leading to expansion

Transposition

Aside from homologous recombination, the overall stability of the genome is interrupted by mobile sequences called transposable elements or transposons. There are different classes of distinct DNA sequences that are able to transport themselves to other locations within the genome. This process utilizes recombination but does not result in an exchange. Rather, a transposon moves directly from one site of the genome to another without an intermediary such as plasmid or phage DNA. This results in rearrangements that create new sequences and change the functions of target sequences. Transposons may be a major source of evolutionary changes in the genome. In some cases they cause disease when inserted into a functioning gene. Three examples are presented below: insertion sequences (IS), transposons (Tn), and retroelements transposing via an RNA intermediate.

A. Insertion sequences (IS) and transposons (Tn)

A characteristic feature of IS transposition is the presence of a pair of short direct repeats of target DNA at either end. The IS itself carries inverted repeats of about 9–13 bp at both ends and depending on the particular class consists of about 750–1500 bp, which contain a single long coding region for transposase (the enzyme responsible for transposition of mobile sequences). Target selection is either random or at particular sites. The presence of inverted terminal repeats and the short direct repeats of host DNA result in a characteristic structure (1). Transposons carry in addition a central region with genetic markers unrelated to transposition, e.g., antibiotic resistance (2). They are flanked either by direct repeats (same direction) or by inverted repeats (opposite direction, 3).



Insertion sequences (IS) and transposons (Tn)

B. Replicative and nonreplicative transposition

With replicative transposition, the original transposon remains in place and creates a new copy of itself, which inserts into a recipient site elsewhere. Thus, this mechanism leads to an increase in the number of copies of the transposon in the genome. This type involves two enzymatic activities: a transposase acting on the ends of the original transposon and resolvase acting on the duplicated copies. In nonreplicative transposition, the transposing element itself moves as a physical entity directly to another site. The donor site is either repaired (in eukaryotes) or may be destroyed (in bacteria) if more than one copy of the chromosome is present.

C. Transposition of retroelements

Retrotransposition requires synthesis of an RNA copy of the inserted retroelement. Retroviruses including the human

immunodeficiency virus and RNA tumor viruses are important retroelements. The first step in retrotransposition is the synthesis of an RNA copy of the inserted retroelement, followed by reverse transcription up to the polyadenylation sequence in the 3′ long terminal repeat (3'LTR). Three important classes of mammalian transposons that undergo or have undergone retrotransposition through an RNA intermediary are shown. Endogenous retroviruses , are sequences that resemble retroviruses but cannot infect new cells and are restricted to one genome. Non viral retrotransposons , lack LTRs and usually other parts of retroviruses. Both types contain reverse transcriptase and are therefore capable of independent transposition. Processed pseudogenes, or retropseudogenes lack reverse transcriptase and cannot transpose independently. They contain two groups: low copy number of processed pseudogenes transcribed by RNA polymerase II and high copy number of mammalian SINE sequences, such as human Alu and the mouse B1 repeat families.

Recombination

Recombination lends the genome flexibility. Without genetic recombination, the genes on each individual chromosome would remain fixed in their particular position. Changes could occur by mutation only, which would be hazardous. Recombination provides the means to achieve extensive restructuring, eliminate unfavorable mutation, maintain and spread favorable mutations, and endow each individual with a unique set of genetic information. This greatly enhances the evolutionary potential of the genome. Recombination must occur between precisely corresponding sequences (homologous recombination) to ensure that not one base pair is lost or added. The newly combined (recombined) stretches of DNA must retain their

original structure in order to function properly. Two types of recombination can be distinguished: (1) generalized or homologous recombination, which in eukaryotes occurs at meiosis and (2) site-specific recombination. A third process, transposition, utilizes recombination to insert one DNA sequence into another without regard to sequence homology. Here we consider homologous recombination, a complex biochemical reaction between two duplexes of DNA. The necessary enzymes, which can involve any pair of homologous sequences, are not considered. Two general models can be distinguished, recombination initiated from a single-strand DNA break and recombination initiated from a double-strand break.

A. Recombination initiated by single-strand breaks

This model assumes that the process starts with breaks at corresponding positions of one of the strands of homologous DNA (same sequences of different parental origin) (1). A nick is made by a single-strand-breaking enzyme (endonuclease) in each molecule at the corresponding site (2), but see below. This allows the free ends of one nicked strand to join with the free ends of the other nicked strand, from the other molecule, to form single-strand exchanges between the two duplex molecules at the recombination joint (3). The recombination joint moves along the duplex (branch migration) (4). This is an important feature because it ensures that sufficient distance for the second nick is present in each of the other strands (5). After the two other strands have joined and gaps have been sealed (6), a reciprocal recombinant molecule is generated (7). Recombination involving DNA duplexes requires topological changes, i.e., either the molecules must be free to rotate or the restraint must be relieved in some other way. This model has an unresolved difficulty: How is it assured that the single-strand nicks

shown in step 2 occur at precisely the same position in the two double helix DNA molecules?

B. Recombination initiated by double-strand breaks

The current model for recombination is based on initial double-strand breaks in one of the two homologous DNA molecules (1). Both strands are cleaved by an endonuclease, and the break is enlarged to a gap by an exonuclease that removes the new 5′ ends of the strands at the break and leaves 3′ single-stranded ends (2). One free 3′ end recombines with a homologous strand of the other molecule (3). This generates a D loop consisting of a displaced strand from the “donor” duplex. The D loop is extended by repair synthesis until the entire gap of the recipient molecule is closed (4). This displaced strand anneals to the single-stranded complementary homologous sequences of the recipient strand and closes the gap (5). DNA repair synthesis from the other 3′ end closes the remaining gap (6). The integrity of the two molecules is restored by two rounds of single strand repair synthesis. In contrast to the single-strand exchange model, the double strand breaks result in hetero-duplex DNA in the entire region that has undergone recombination. An apparent disadvantage is the temporary loss of information in the gaps after the initial cleavage. However, the ability to retrieve this information by resynthesis from the other duplex avoids permanent loss.

DNA Polymorphism

Genetic polymorphism is the existence of variants with respect to a gene locus (alleles), a chromosome structure(e.g., sizeofcentromeric heterochromatin), a gene product (variants in enzymatic activity or binding affinity), or a phenotype. The term DNA polymorphism refers to a wide range of variations in nucleotide base composition, length of nucleotide repeats, or single nucleotide variants. DNA polymorphisms are important as genetic markers to identify and distinguish alleles at agene locus and to determine their parental origin.

A. Single nucleotide polymorphism (SNP)

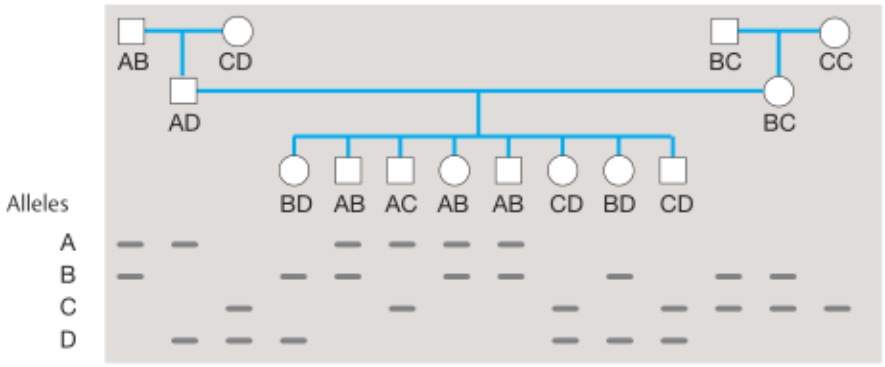
These allelic variants differ in a single nucleotide at a specific position. At least one in a thousand DNA bases differs among individuals (1). The detection of SNPs does not require gel electrophoresis. This facilitates large-scale detection. A SNP can be visualized in a Southern blot as a restriction fragment length polymorphism (RFLP) if the difference in the two alleles corresponds to a difference in the recognition site of a restriction enzyme (see Southern blot, p. 62). B. Simple sequence length polymorphism (SSLP) These allelic variants differ in the number of tandemly repeated short nucleotide sequences in noncoding DNA. Short tandem repeats (STRs) consist of units of 1,2,3,or 4 base pairs repeated from 3 to about 10 times. Typical short tandem repeats are CA repeats in the 5′ to 3′ strand, i.e., alternating CG and AT base pairs in the double strand. Each allele is defined by the number of CA repeats, e.g., 3 and 5, as shown (1). These are also called microsatellites. The size differences due to the number of repeats are determined by PCR. Variable number of tandem repeats (VNTR), also called minisatellites, consist of repeat units of 20–200 base pairs

(2). C. Detection of SNP by oligonucleotide hybridization analysis Oligonucleotides, short stretches of about 20 nucleotides with a complementary sequence to the single-stranded DNA to be examined, will hybridize completely only if perfectly matched. If there is a difference of even one base, such as due to an SNP, the resulting mismatch can be detected because the DNA hybrid is unstable and gives no signal.

D. Detection of STRs by PCR

Short tandem repeats (STRs) can be detected by the polymerase chain reaction (PCR). The allelic regions of a stretch of DNA are amplified; the resulting DNA fragments of different sizes are subjected to electrophoresis; and their sizes are determined.

E. CEPH families



Alleles A, B, C, D at a marker locus show all possible pairwise combinations

CEPH family

An important step in gene identification is the analysis of large families by linkage analysis of polymorphic marker loci on a specific chromosomal region near a locus of interest. Large families are of particular value. DNA from such families has been collected by the Centre pour l'Étude du Polymorphisme Humain (CEPH) in Paris, now called the Centre Jean Dausset, after the founder. Immortalized cell lines are stored from each family. A CEPH family consists of four grandparents, the two parents, and eight children. If four alleles are present

at a given locus they are designated A, B, C, and D. Starting with the grandparents, the inheritance of each allele through the parents to the grandchildren can be traced (shown here as a schematic pattern in a Southern blot). Of the four grandparents shown, three are heterozygous (AB, CD, BC) and one is homozygous (CC). Since the parents are heterozygous for different alleles (AD the father and BC the mother), all eight children are heterozygous (BD, AB, AC, or CD).

Mutation Due to Different Base Modifications

Mutations can result from chemical or physical events that lead to base modification. When they affect the base-pairing pattern, they interfere with replication or transcription. Chemical
substancees able to induce such changes are called mutagens. Mutagens cause mutations in different ways. Spontaneous oxidation, hydrolysis, uncontrolled methylation, alkylation, and ultraviolet irradiation result in alterations that modify nucleotide bases. DNA-reactive chemicals change nucleotide bases into different chemical structures or remove a base.

A. Deamination and methylation

Cytosine, adenine, and guanine contain an amino group. When this is removed (deamination), a modified base with a different basepairing pattern results. Nitrous acid typically removes the amino group. This also occurs spontaneously at a rate of 100 bases per genome per day (Alberts et al., 1994, p. 245). Deamination of cytosine removes the amino group in position 4 (1). The resulting molecule is uracil (2). This

pairs with adenine rather than guanine. Normally this change is efficiently repaired by uracil-DNA glycosylase. Deamination at the RNA level occurs in RNA editing (see Expression of genes). Methylation of the carbon atom in position 5 of cytosine results in 5 methylcytosine, containing a methyl group in position 5 (3). Deamination of 5-methylcytosine will result in a change to thymine, containing an oxygen in position 4 instead of an amino group (4). This mutation will not be corrected because thymine is a natural base. Adenine (5) can be deaminated in position 6 to form hypoxanthine, which contains an oxygen in this position instead of an amino group (6), and which pairs with cytosine instead of thymine. The resulting change after DNA replication is a cytosine instead of a thymine in the mutant strand.

B. Depurination

About 5000 purine bases (adenine and guanine) are lost per day from DNA in each cell (depurination) owing to thermal fluctuations. Depurination of DNA involves hydrolytic cleavage of the N-glycosyl linkage of deoxyribose to the guanine nitrogen in position 9.This leaves a depurinated sugar. The loss of a base pair will lead to a deletion after the next replication if not repaired in time (see DNA repair).

C. Alkylation

Alkylation is the introduction of a methyl or an ethyl group into a molecule. The alkylation of guanine involves the replacement of the hydrogen bond to the oxygen atom in position 6 by a methyl group, to form 6-methylguanine. This can no longer pair with cytosine. Instead, it will pair with thymine. Thus, after the next replication the opposite cytosine (C) is replaced by a thymine (T) in the mutant daughter molecule. Important alkylating agents are ethylnitroso urea (ENU), ethylmethane sulfonate (EMS),

dimethylnitrosamine, and N-methyl-N-nitroN-nitrosoguanidine.

D. Nucleotide base analogue

Base analogs are purines or pyrimidines that are similar enough to the regular nucleotide DNA bases to be incorporated into the new strand during replication. 5-Bromodeoxyuridine (5BrdU) is an analog of thymine. It contains a bromine atom instead of the methyl group in position 5.Thus, it can be incorporated into the new DNA strand during replication. However, the presence of the bromine atom causes ambiguous and often wrong base pairing. E. UV-light-induced thymine dimers Ultraviolet irradiation at 260 nm wavelength induces covalent bonds between adjacent thymine residues at carbon positions 5 and 6. If located within a gene, this will interfere with replication and transcription unless repaired. Another important type of UV-induced change is a photoproduct consisting of a covalent bond between the carbons in positions 4 and 6 of two adjacent nucleotides, the 4–6 photoproduct.

Changes in DNA (Mutations)

When it was recognized that changes (mutations) in genes occur spontaneously (T. H. Morgan, 1910) and can be induced by X-rays (H. J. Muller, 1927), the mutation theory of heredity became a cornerstone of early genetics. Genes were defined as mutable units, but the question what genes and mutations are remained. Today we know that mutations are changes in the structure of DNA and their functional consequences. The study of mutations is important for several reasons. Mutations cause diseases, including all forms of cancer. They can be induced by chemicals and by irradiation. Thus, they represent a link between heredity and environment. And without mutations, well-

organized forms of life would not have evolved.

The following two plates summarize the chemical nature of mutations.

A. Error in replication

The synthesis of a new strand of DNA occurs by semiconservative replication based on complementary base pairing. Errors in replication occur at a rate of about 1 in 10 5 . This rate is reduced to about 1 in 10 7 to 10 9 by proofreading mechanisms. When an error in replication occurs before the next cell division (here referred to as the first division after the mutation), e.g., a cytosine (C) might be incorporated instead of an adenine (A) at the fifth base pair as shown here, the resulting mismatch will be recognized and eliminated by mismatchre pair in most cases.

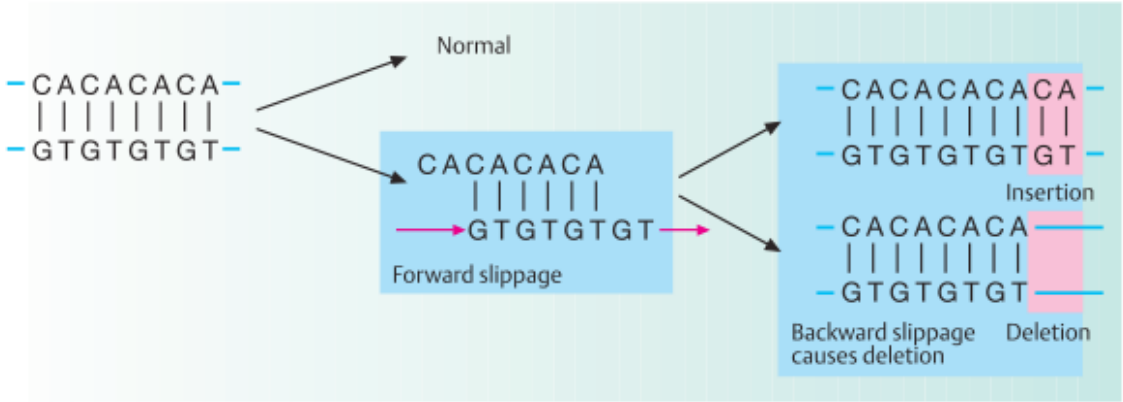
However, if the error is undetected and allowed tostand, the next(second) division will result in a mutant molecule containing a CG instead of an AT pair at this position. This mutation will be perpetuated in all daughter cells. Depending on its location within or outside of the coding region of a gene, functional consequences due to a change in a codon could result.

B. Mutagenic alteration of a nucleotide

A mutation may result when a structural change of a nucleotide affects its base-pairing capability. The altered nucleotide is usually present in one strand of the parent molecule. If this leads to incorporation of a wrong base, such as a C instead of a T in the fifth base pair as shown here, the next (second) round of replication will result in two mutant molecules.

C. Replication slippage

A different class of mutations does not involve an alteration of individual nucleotides, but results from incorrect alignment between allelic or nonallelic DNA sequences during replication. When the template strand contains short tandem repeats, e.g., CA repeats as in microsatellites (see DNA polymorphism and Part II, Genomics), the newly replicated strand and the template strand may shift their positions relative to each other. With replication or polymerase slippage, leading to incorrect pairing of repeats, some repeats are copied twice or not at all, depending on the direction of the shift. One can distinguish forward slippage (shown here) and backward slippage with respect to the newly replicated strand. If the newly synthesized DNA strand slips forward, a region of nonpairing remains in the parental strand. Forward slippage results in an insertion. Backward slippage of the new strand results in deletion. Microsatellite instability is a characteristic feature of hereditary nonpolyposis cancer of the colon (HNPCC). HNPCC genes are localized on human chromosomes at 2p15–22 and 3p21.3. About 15% of all colorectal, gastric, and endometrial carcinomas show microsatellite instability. Replication slippage must be distinguished from unequal crossing-over during meiosis. This is the result of recombination between adjacent, but not allelic, sequences on nonsister chromatids of homologous chromosomes (Figures redrawn from Brown, 1999).



Replication slippage