Spontaneous mutations

Spontaneous mutations arise from a variety of sources, including errors in DNA replication, spontaneous lesions, and transposable genetic elements. The first two are considered in this section

Errors in DNA replication

An error in DNA replication can occur when an illegitimate nucleotide pair (say, A–C) forms in DNA synthesis, leading to a base substitution.

Each of the bases in DNA can appear in one of several forms, called **tautomers,** which are isomers that differ in the positions of their atoms and in the bonds between the atoms. The forms are in equilibrium. The **keto** form of each base is normally present in DNA (Figure 16-1), whereas the **imino** and **enol** forms of the bases are rare. The ability of the wrong tautomer of one of the standard bases to mispair and cause a mutation in the course of DNA replication was first noted by Watson and Crick when they formulated their model for the structure of DNA (Chapter 8). Figure 16-2 demonstrates some possible mispairs resulting from the change of one tautomer into another, termed a tautomeric shift.

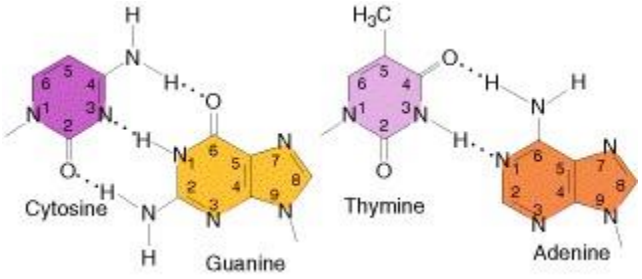


Figure 16-1

Pairing between the normal (keto) forms of the bases.

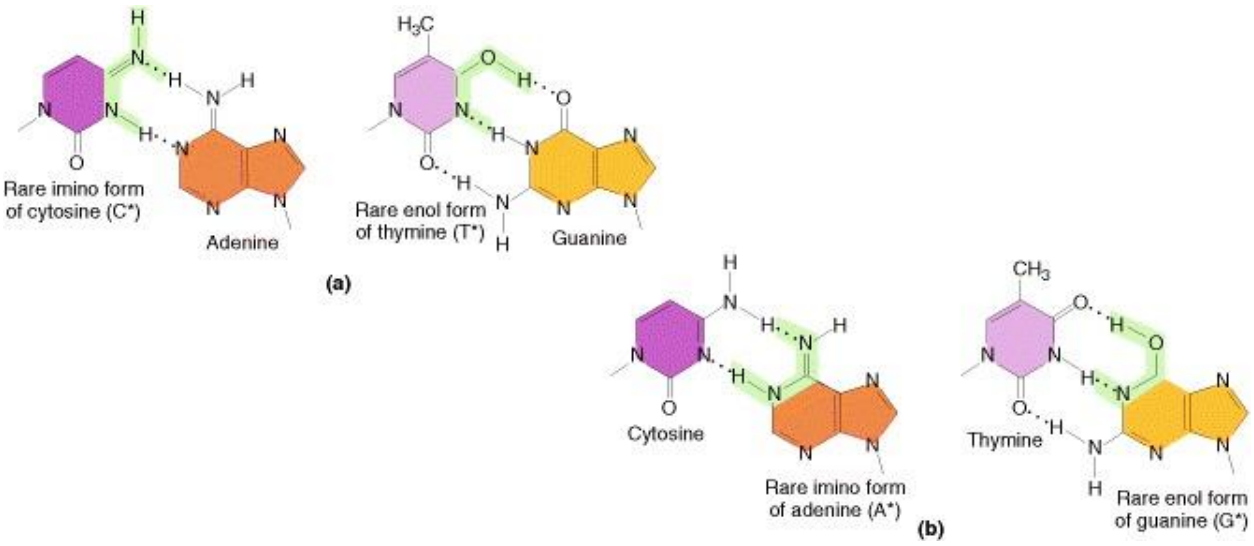


Figure 16-2

Mismatched bases. (a) Mispairs resulting from rare tautomeric forms of the pyrimidines; (b) mispairs resulting from rare tautomeric forms of the purines.

Mispairs can also result when one of the bases becomes **ionized.** This type of mispair may occur more frequently than mispairs due to imino and enol forms of bases.

Transitions.

All the mispairs described so far lead to **transition mutations,** in which a purine substitutes for a purine or a pyrimidine for a pyrimidine (Figure 16-3). The bacterial DNA polymerase III (Chapter 8) has an editing capacity that recognizes such mismatches and excises them, thus greatly reducing the observed mutations. Another repair system (described later in this chapter) corrects many of the mismatched bases that escape correction by the polymerase editing function.

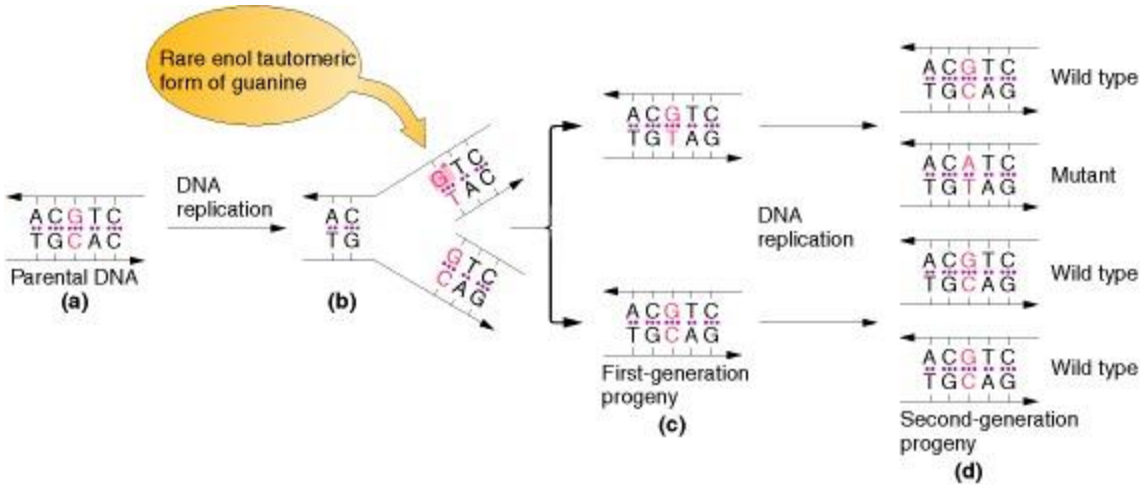


Figure 16-3

Figure 16-3

Mutation by tautomeric shifts in the bases of DNA. (a) In the example diagrammed, a guanine undergoes a tautomeric shift to its rare enol form (G*) at the time of replication. (b) In its enol form, it pairs with thymine. (c and d) In the next replication, the guanine shifts back to its more stable keto form. The thymine incorporated opposite the enol form of guanine, seen in part b, directs the incorporation of adenine in the subsequent replication, shown in parts c and d. The net result is a GC → AT mutation. If a guanine undergoes a tautomeric shift from the common keto form to the rare enol form at the time of incorporation (as a nucleoside triphosphate, rather than in the template strand diagrammed here), it will be incorporated opposite thymine in the template strand and cause an AT → GC mutation. (From E. J. Gardner and D. P. Snustad, *Principles of Genetics,* 5th ed. Copyright © 1984 by John Wiley & Sons, New York.)

Transversions.

In **transversion mutations,** a pyrimidine substitutes for a purine or vice versa. Transversions cannot be generated by the mismatches depicted in Figure 16-2. With bases in the DNA in the normal orientation, creation of a transversion by a replication error would require, at some point in the course of replication, mispairing of a purine with a purine or a pyrimidine with a pyrimidine. Although the dimensions of the DNA double helix render such mispairs energetically unfavorable, we now know from X-ray

diffraction studies that G–A pairs, as well as other purine–purine pairs, can form.

Frameshift mutations.

Replication errors can also lead to **frameshift mutations.** Recall from Chapter 10 that such mutations result in greatly altered proteins.

In the mid-1960s, George Streisinger and his coworkers deduced the nucleotide sequence surrounding different sites of frameshift mutations in the lysozyme gene of phage T4. They found that these mutations often occurred at repeated sequences and formulated a model to account for frameshifts in DNA synthesis. In the Streisinger model (Figure 16-4), frameshifts arise when loops in single-stranded regions are stabilized by the "slipped mispairing" of repeated sequences.

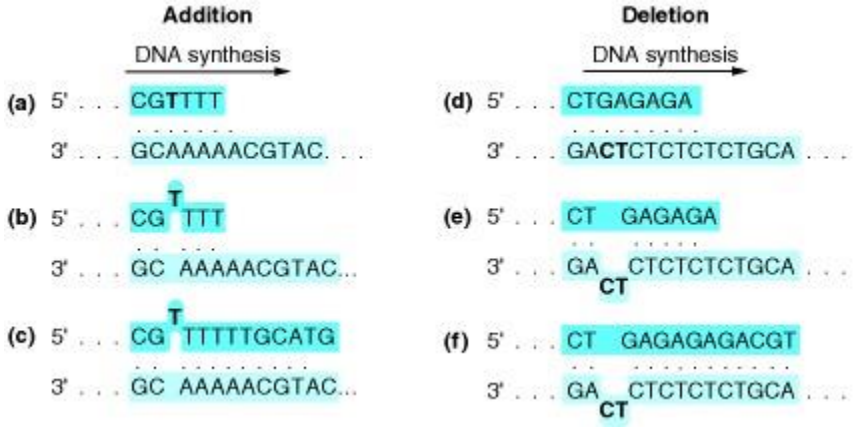


Figure 16-4

A simplified version of the Streisinger model for frameshift formation. (a–c) In DNA synthesis, the newly synthesized strand slips, looping out one or several bases. This loop is stabilized by the pairing afforded by the repetitive-sequence unit (the A bases in this case). An addition of one base pair, A–T, will result at the next round of replication in this example. (d–f) If, instead of the newly synthesized strand, the template strand slips, then a deletion results. Here the repeating unit is a CT dinucleotide. After slippage, a deletion of two base pairs (C–G and T–A) would result at the next round of replication.

In the 1970s, Jeffrey Miller and his co-workers examined mutational hot spots in the *lacI* gene of *E. coli.* As already mentioned, hot spots are sites in a

gene that are much more mutable than other sites. The *lacI* work showed that certain hot spots result from repeated sequences, just as predicted by the Streisinger model. Figure 16-5 depicts the distribution of spontaneous mutations in the *lacI* gene. Compare this distribution with that in the *rII* genes of T4 seen by Benzer (see Figure 9-26). Note how one or two mutational sites dominate the distribution in both cases. In *lacI*, a four-base-pair sequence repeated three times in tandem in the wild type is the cause of the hot spots (for simplicity, only one strand of the double strand of DNA is indicated):

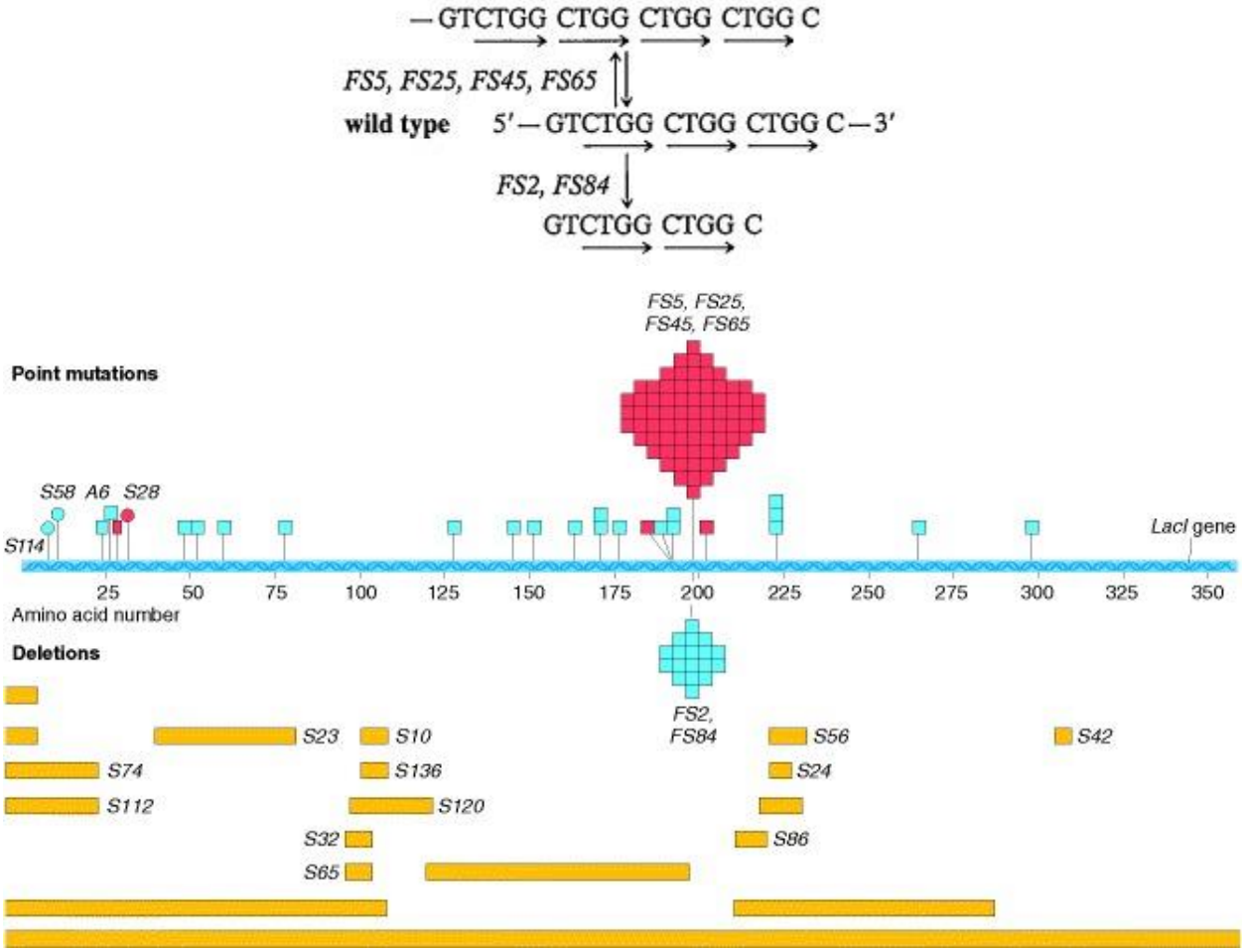


Figure 16-5

The distribution of 140 spontaneous mutations in *lacI*. Each occurrence of a point mutation is indicated by a box. Red boxes designate fast-reverting mutations. Deletions (gold) are represented below. The *I* map is given in terms of the amino acid number in the corresponding *I*-encoded *lac* repressor. Allele numbers refer to mutations that have been analyzed at

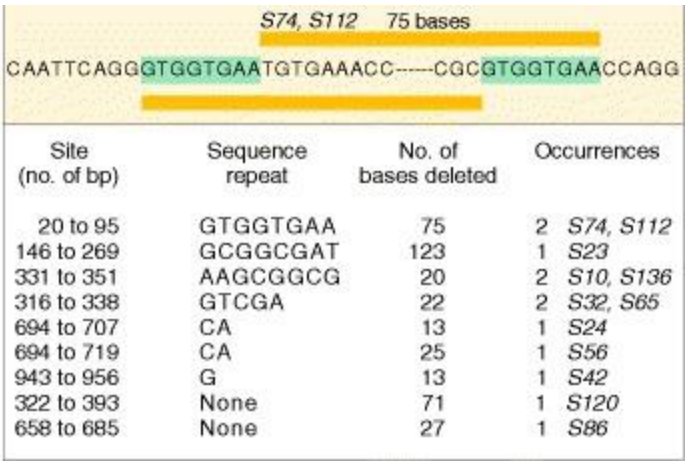
the DNA sequence level. The mutations *S114* and *S58* (circles) result from the insertion of transposable elements (see Chapter 20). *S28* (red circle) is a duplication of 88 base pairs. (From P. J. Farabaugh, U. Schmeissner, M. Hofer, and J. H. Miller, *Journal of Molecular Biology* 126, 1978, 847.)

The major hot spot, represented here by the mutations *FS5*, *FS25*, *FS45*, and *FS65*, results from the addition of one extra set of the four bases CTGG to one strand of the DNA. This hot spot reverts at a high rate, losing the extra set of four bases. The minor hot spot, represented here by the mutations *FS2* and *FS84*, results from the loss of one set of the four bases CTGG. This mutant does not readily regain the lost set of four base pairs.

How can the Streisinger model explain these observations? The model predicts that the frequency of a particular frameshift depends on the number of base pairs that can form during the slipped mispairing of repeated sequences. The wild-type sequence shown for the *lacI* gene can slip out one CTGG sequence and stabilize structure by forming nine base pairs. (Can you work this out by applying the model in Figure 16-4 to the sequence shown for *lacI*?) Whether a deletion or an addition is generated depends on whether the slippage occurs on the template or on the newly synthesized strand, respectively. In a similar fashion, the addition mutant can slip out one CTGG sequence and stabilize a structure with 13 base pairs (verify this for the *FS5* sequence shown for *lacI*), which explains the rapid reversion of mutations such as *FS5*. However, there are only five base pairs available to stabilize a slipped-out CTGG in the deletion mutant, accounting for the infrequent reversion of mutations such as *FS2* in the sequence shown for*lacI*.

Deletions and duplications.

Large **deletions** (more than a few base pairs) constitute a sizable fraction of sponta-neous mutations, as shown in Figure 16-5. The majority, although not all, of the deletions occur at repeated sequences. Figure 16-6 shows the results for the first 12 deletions analyzed at the DNA sequence level, presented by Miller and his co-workers in 1978. Further studies showed that hot spots for deletions are in the longest repeated sequences. **Duplications** of segments of DNA have been observed in many organisms. Like deletions, they often occur at sequence repeats.



Site (no. of bp)	Sequence repeat	No. of bases deleted	Occurrences
20 to 95	GTGGTGAA	75	2 *S74, S112*
146 to 269	GCGGCGAT	123	1 *S23*
331 to 351	AAGCGGCG	20	2 *S10, S136*
316 to 338	GTCGA	22	2 *S32, S65*
694 to 707	CA	13	1 *S24*
694 to 719	CA	25	1 *S56*
943 to 956	G	13	1 *S42*
322 to 393	None	71	1 *S120*
658 to 685	None	27	1 *S86*

Figure 16-6

Deletions in *lacI.* Deletions occurring in *S74* and *S112* are shown at the top of the figure. As indicated by the gold bars, one of the sequence repeats (aqua) and all the intervening DNA are deleted, leaving one copy of the repeated sequence. All mutations were analyzed by direct DNA sequence determination. (From P. J. Farabaugh, U. Schmeissner, M. Hofer, and J. H. Miller, *Journal of Molecular Biology*126, 1978, 847.)

How do deletions and duplications form? Several mechanisms could account for their formation. Deletions may be generated as replication errors. For example, an extension of the Streisinger model of slipped mispairing (Figure 16-4) could explain why deletions predominate at short repeated sequences. Alternatively, deletions and duplications could be generated by recombinational mechanisms (to be described in Chapter 19).

Spontaneous lesions

In addition to replication errors, **spontaneous lesions,** naturally occurring damage to the DNA, can generate mutations. Two of the most frequent spontaneous lesions result from depurination and deamination.

Depurination, the more common of the two, consists of the interruption of the glycosidic bond between the base and deoxyribose and the subsequent loss of a guanine or an adenine residue from the DNA (Figure 16-

7). A mammalian cell spontaneously loses about 10,000 purines from its DNA in a 20-hour cell-generation period at 37°C. If these lesions were to persist, they would result in significant genetic damage because, in replication, the resulting **apurinic sites** cannot specify a base complementary to the original purine. However, as we shall see later in the chapter, efficient repair systems remove apurinic sites. Under certain conditions (to be described later), a base can be inserted across from an apurinic site; this insertion will frequently result in a mutation.

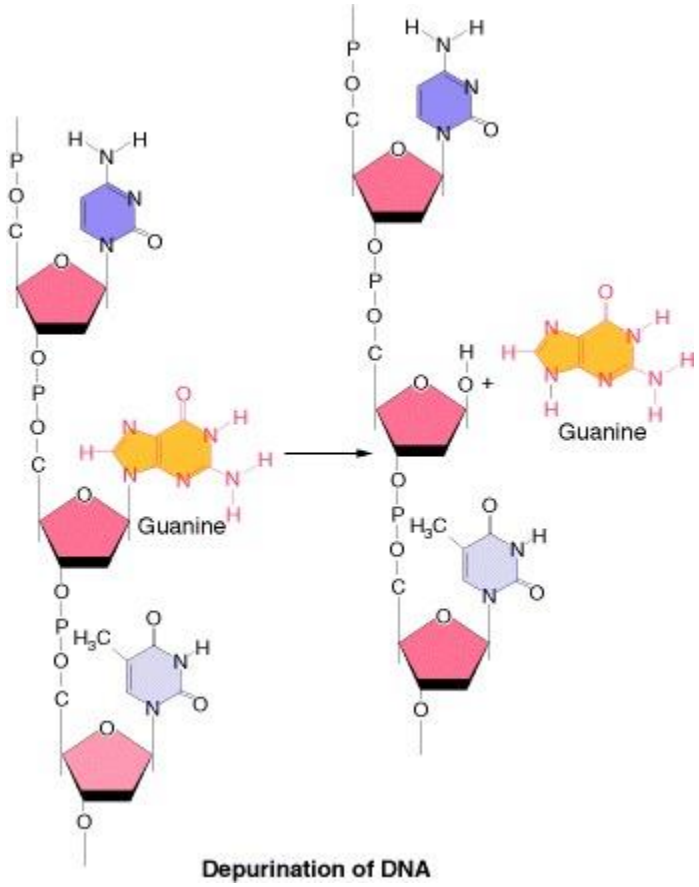


Figure 16-7

The loss of a purine residue (guanine) from a single strand of DNA. The sugar-phosphate backbone is left intact.

The **deamination** of cytosine yields uracil (Figure 16-8a). Unrepaired uracil residues will pair with adenine in replication, resulting in the conversion of a G–C pair into an A–T pair (a **GC → AT transition**). In 1978, deaminations at certain cytosine residues were found to be the cause of one type of mutational hot spot. DNA sequence analysis of GC → AT transition hot spots

in the *lacI* gene showed that 5-methylcytosine residues are pres-ent at each hot spot. (Certain bases in prokaryotes and eukaryotes are methylated.) Some of the data from this *lacI*study are shown in Figure 16-9. The height of each bar on the graph represents the frequency of mutations at each of a number of sites. It can be seen that the positions of 5-methylcytosine residues correlate nicely with the most mutable sites.

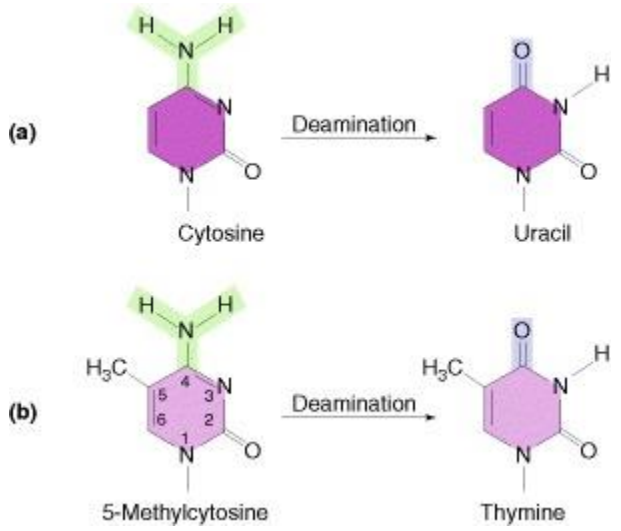


Figure 16-8

Deamination of (a) cytosine and (b) 5-methylcytosine.

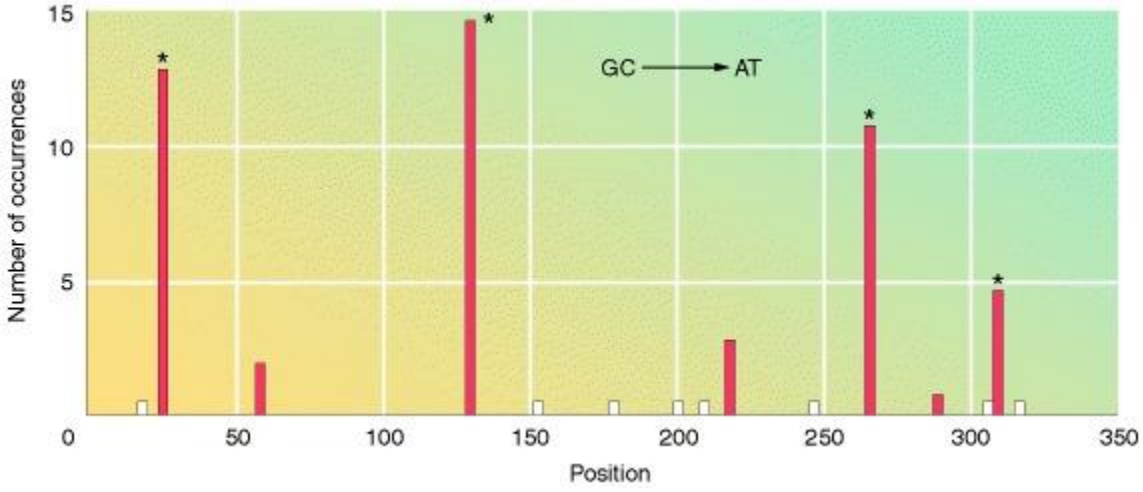


Figure 16-9

5-Methylcytosine hot spots in *E. coli.* Nonsense mutations occurring at 15 different sites in *lacI* were scored. All result from the GC → AT transition. The asterisks (*) mark the positions of 5-methylcytosines. Open bars depict sites at which the GC → AT change could be detected but at which no

mutations occurred in this particular collection. It can be seen that 5-methylcytosine residues are hot spots for the GC → AT transition. Of 50 independently occurring mutations, 44 were at the 4 methylated cytosine sites and only 6 were at the 11 unmethylated cytosines. (From C. Coulondre et al., Nature 274, 1978, 775.).

Why are 5-methylcytosines hot spots for mutations? One of the repair enzymes in the cell, uracil-DNA glycosylase, recognizes the uracil residues in the DNA that arise from deaminations and excises them, leaving a gap that is subsequently filled (a process to be described later in the chapter). However, the deamination of 5-methylcytosine (Figure 16-8b) generates thymine (5-methyluracil), which is not recognized by the enzyme uracil-DNA glycosylase and thus is not repaired. Therefore, C → T transitions generated by deamination are seen more frequently at 5-methylcytosine sites, because they escape this repair system.

A consequence of the frequent mutation of 5-methylcytosine to thymine is the underrepresentation of CpG dinucleotides in higher cells, because this sequence is methylated to give 5-methyl-CpG, which is gradually converted into TpG.

Oxidatively damaged bases represent a third type of spontaneous lesion implicated in mutagenesis. Active oxygen species, such as superoxide radicals ($O_2\cdot$), hydrogen peroxide (H_2O_2), and hydroxyl radicals ($OH\cdot$), are produced as by-products of normal aerobic metabolism. They can cause oxidative damage to DNA, as well as to precursors of DNA (such as GTP), which results in mutation and which has been implicated in a number of human diseases. Figure 16-10 shows two products of oxidative damage. The 8-oxo-7-hydrodeoxyguanosine (8-oxodG, or GO) product frequently mispairs with A, resulting in a high level of G → T transversions. Thymidine glycol blocks DNA replication if unrepaired but has not yet been implicated in mutagenesis.

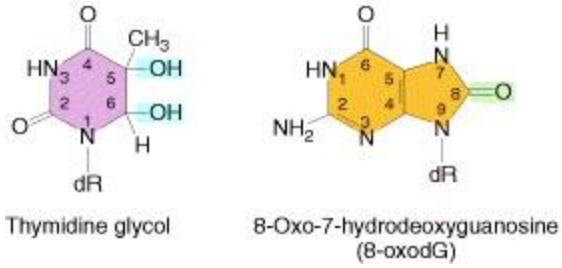


Figure 16-10

DNA damage products formed after attack by oxygen radicals. dR = deoxyribose.

MESSAGE

Spontaneous mutations can be generated by different processes. Replication errors and spontaneous lesions generate most of the base-substitution and frameshift mutations. Replication errors may also cause some deletions that occur in the absence of mutagenic treatment.

Spontaneous mutations and human diseases

DNA sequence analysis has revealed the mutations responsible for a number of human hereditary diseases. The previously discussed studies of bacterial mutations allow us to suggest mechanisms that cause these human disorders.

A number of these disorders are due to **deletions** or **duplications** involving repeated sequences. For example, mitochondrial encephalomyopathies are a group of disorders affecting the central nervous system or the muscles (Kearns-Sayre syndrome). They are characterized by dysfunction of oxidation phosphorylation (a function of the mitochondria) and by changes in mitochondrial structure. These disorders have been shown to result from deletions that occur between repeated sequences. Figure 16-11 depicts one of these deletions. Note how similar it is in form to the spontaneous *E. coli* deletions shown in Figure 16-6.

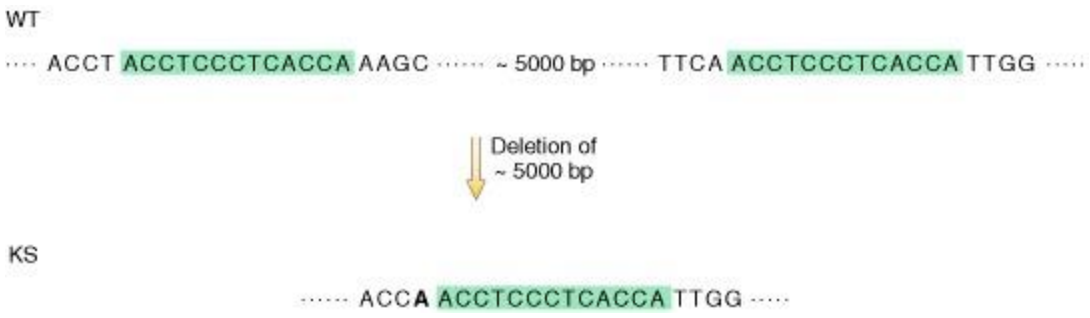


Figure 16-11

Sequences of wild-type (WT) mitochondrial DNA and deleted DNA (KS) from a patient with Kearns-Sayre syndrome. The 13-base boxed sequence is identical in both WT and KS and serves as a breakpoint for the DNA deletion. A single base (boldface type) is altered in KS, aside from the deleted segment.

A common mechanism that is responsible for a number of genetic diseases is the **expansion of a three-base-pair repeat,** as in fragile X syndrome (Figure 16-12). This syndrome is the most common form of inherited mental retardation, occurring in close to 1 of 1500 males and 1 of 2500 females. It is evidenced cytologically by a fragile site in the X chromosome that results in a break in vitro.

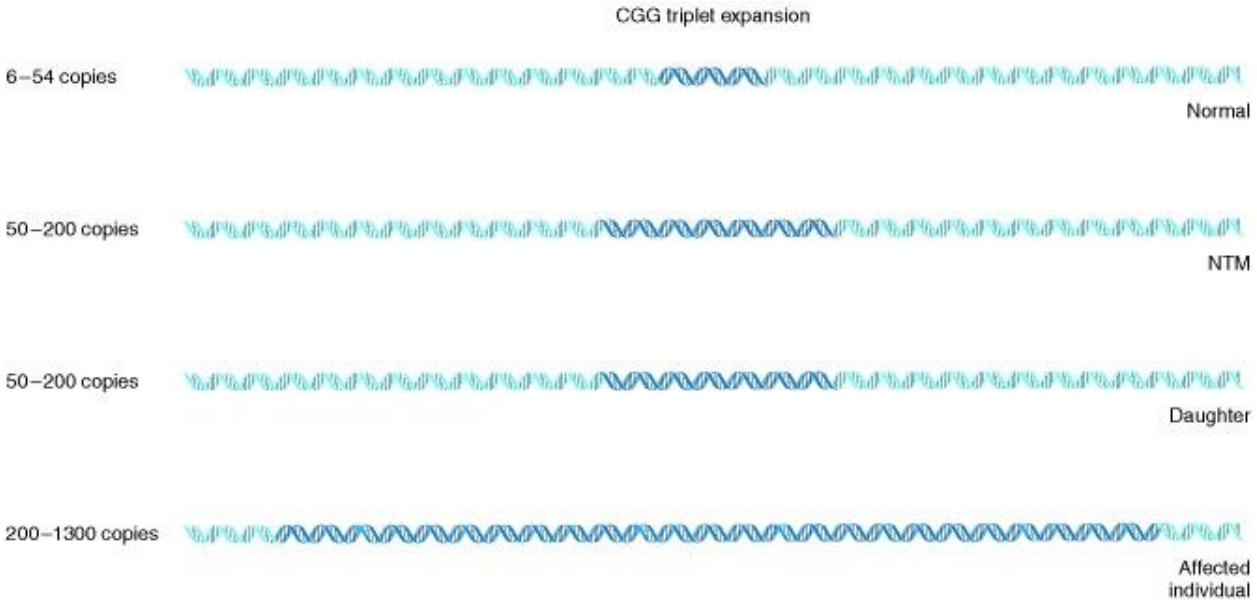


Figure 16-12

Expansion of the CGG triplet in the *FMR-1* gene seen in the fragile X syndrome. Normal persons have from 6 to 54 copies of the CGG repeat, whereas members of susceptible families display an increase (premutation) in the number of repeats: normally transmitting males (NTMs) and their daughters are phenotypically normal but display from 50 to 200 copies of the CGG triplet; the number of repeats expands to some 200 to 1300 in those showing full symptoms of the disease.

The inheritance of fragile X syndrome is unusual in that 20 percent of the males with a fragile X chromosome are phenotypically normal but transmit the affected chromosome to their daughters, who also appear normal. These males are said to be normally transmitting males (NTMs). However, the sons of the daughters of the NTMs frequently display symptoms. The fragile X syndrome results from mutations in a $(CGG)_n$ repeat in the coding sequence of the *FMR-1* gene. Patients with the disease show specific methylation, induced by the mutation, at a nearby CpG cluster, resulting in reduced *FMR-1* expression.

Why do symptoms develop in some persons with a fragile X chromosome and not in others? The answer seems to lie in the number of CGG repeats in the *FMR-1* gene. Humans normally show a considerable variation in the number of CGG repeats in the *FMR-1* gene, ranging from 6 to 54, with 29 repeats in the most frequent allele. [The variation in CGG repeats produces a corresponding variation in the number of arginine residues (CGG is an arginine codon) in the *FMR-1*-encoded protein.] Both NTMs and their daughters have a much larger number of repeats, ranging from 50 to 200. These increased repeats have been termed **premutations.** All premutation alleles are unstable. The males and females with symptoms of the disease, as well as many carrier females, have additional insertions of DNA, suggesting repeat numbers of 200 to 1300. The frequency of expansion has been shown to increase with the size of the DNA insertion (and thus, presumably, with the number of repeats). Apparently, the number of repeats in the premutation alleles found in NTMs and their daughters is above a certain threshold and thus is much more likely to expand to a full mutation than is the case for normal persons.

The proposed mechanism for these repeats is a slipped mispairing in DNA synthesis (as shown in Figure 16-6) involving a one-step expansion of the four-base-pair sequence CTGG. However, the extraordinarily high frequency of mutation at the three-base-pair repeats in the fragile X syndrome suggests that in human cells, after a threshold level of about 50 repeats, the replication machinery cannot faithfully replicate the correct sequence, and large variations in repeat numbers result.

A second inherited disease, X-linked spinal and bulbar muscular atrophy (known as Kennedy disease), also results from the amplification of a three-base-pair repeat, in this case a repeat of the CAG triplet. Kennedy disease, which is characterized by progressive muscle weakness and atrophy, results from mutations in the gene that encodes the androgen receptor. Normal persons have an average of 21 CAG repeats in this gene, whereas affected patients have repeats ranging from 40 to 52.

Myotonic dystrophy, the most common form of adult muscular dystrophy, is yet another example of sequence expansion causing a human disease. Susceptible families display an increase in severity of the disease in successive generations; this increase is caused by the progressive amplification of a CTG triplet at the 3′ end of a transcript. Normal people possess, on average, five copies of the CTG repeat; mildly affected people have approximately 50 copies, and severely affected people have more than 1000 repeats of the CTG triplet. Additional examples of triplet expansion are still appearing—for instance, Huntington disease.