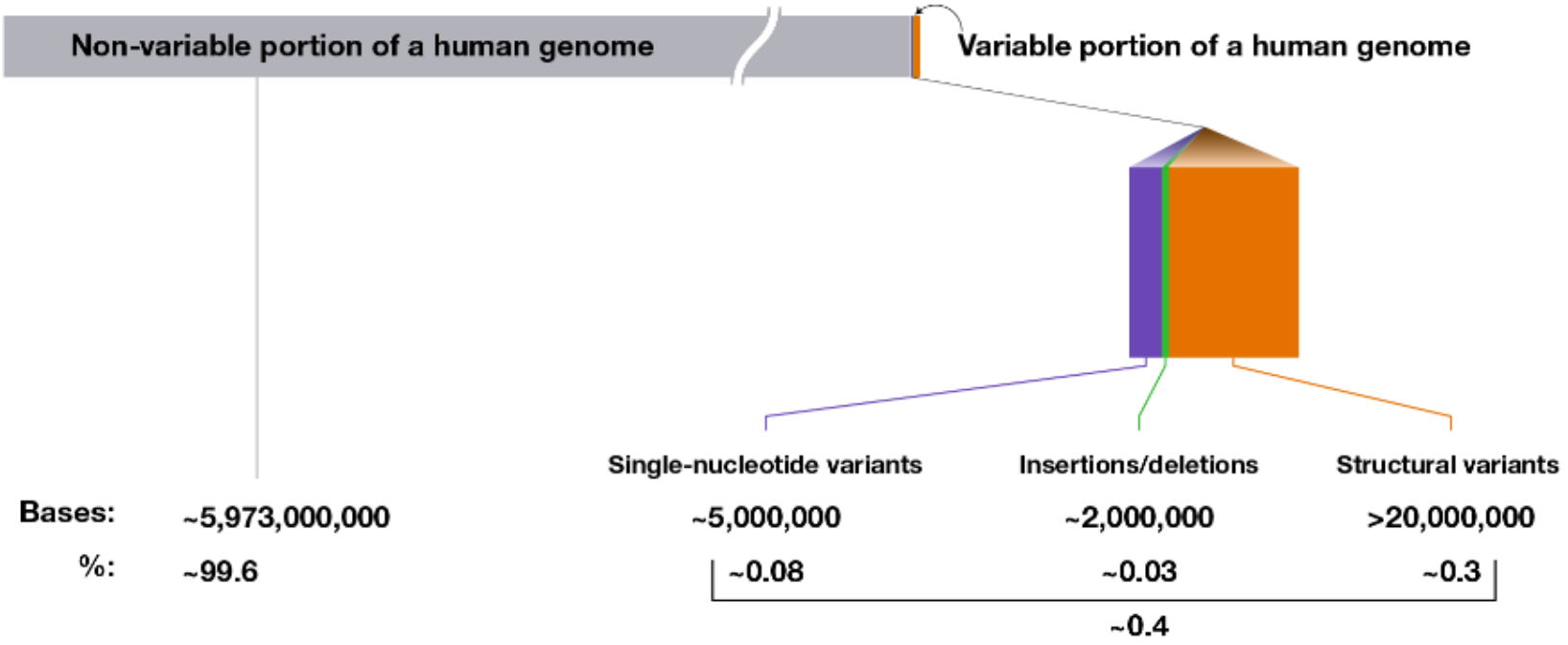
Genetic Variation I

- Describe different types of genetic variations
- Give examples of genetic variations

Similarity of DNA

- One copy of the human genome contains about 3 billion nucleotides (in 23 chromosomes)
- Two random people are 99.6% similar



The Nature of Mutations

- A **mutation** is a change in a DNA sequence (is rare in a population and typically affects the phenotype)
  - Relatively rare, with a population frequency less than 1%.
  - They occur both in the nuclear and in mitochondrial genomes.
- A **polymorphism** ("many forms," describing multiple alleles at a locus): is a genetic change that (is frequent in a population >1%
- The term mutant refers to the phenotype
- An agent that causes a mutation is called a mutagen

GJB2:c.109G>A

chr13-20763612 C>T | p.Val37Ile | NM_004004.6 |

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
East Asian	1665	19952	96	0.08345
Ashkenazi Jewish	83	10342	0	0.008026
Other	31	7212	0	0.004298
Latino/Admixed American	95	35428	1	0.002681
European (Finnish)	42	25104	0	0.001673
European (non-Finnish)	179	128578	1	0.001392
African/African-American	25	24964	1	0.001001
South Asian	12	30584	0	0.0003924
XX	1083	129104	53	0.008389
XY	1049	153060	46	0.006854
Total	**2132**	**282164**	**99**	**0.007556**

Mutation, polymorphism and variant

- “A **mutation** is defined as a permanent change in the nucleotide sequence with a frequency below 1%
- A **polymorphism** is defined as a variant with a frequency above 1%
- The terms “mutation” and “polymorphism,” however, which have been used widely, often lead to confusion because of incorrect assumptions of pathogenic and benign effects, respectively.
- Thus, it is recommended that both terms be replaced by the term “variant”” ACMG 2015 guidelines

Types of Mutations

- Human DNA variants can be classified as large scale versus small scale, common versus rare, and pathogenic versus nonpathogenic.
- Human genetic variation ranges from single nucleotide changes through to gains or losses of whole chromosomes.
- Small-scale variants normally have their primary effect, if they have any effect, on a single gene, whereas large-scale variants usually affect several or many genes.

All mutations fall into two basic categories:

- Those that produce changes in a single gene are known as **gene mutations.**
- Those that produce changes in whole chromosomes are known as **chromosomal mutations.**

Gene mutations

- Gene mutations, including base pair substitutions, insertions, and deletions, can originate by either of two basic mechanisms:
  - Errors introduced during the normal process of DNA replication, or
  - Mutations arising from a failure to repair DNA after damage and to return its sequence to what it was before the damage.

DNA DAMAGE AND REPAIR MECHANISMS

- Some DNA variants arise from errors in DNA replication or recombination, but a major source is a failure to repair DNA damage.
- Chemical reactions and some physical processes constantly damage genomic DNA.
  - The majority are corrected using the undamaged strand as a template.
  - Some base changes escape repair, and an incorrect base serves as a template in replication.
  - The human genome contains genes for > 130 repair proteins: more than 99.9% of errors of DNA replication are corrected.
- The agents that damage DNA can either be external to the cell or they may arise as undesired effects of internal cell chemistry.

Causes of gene mutations

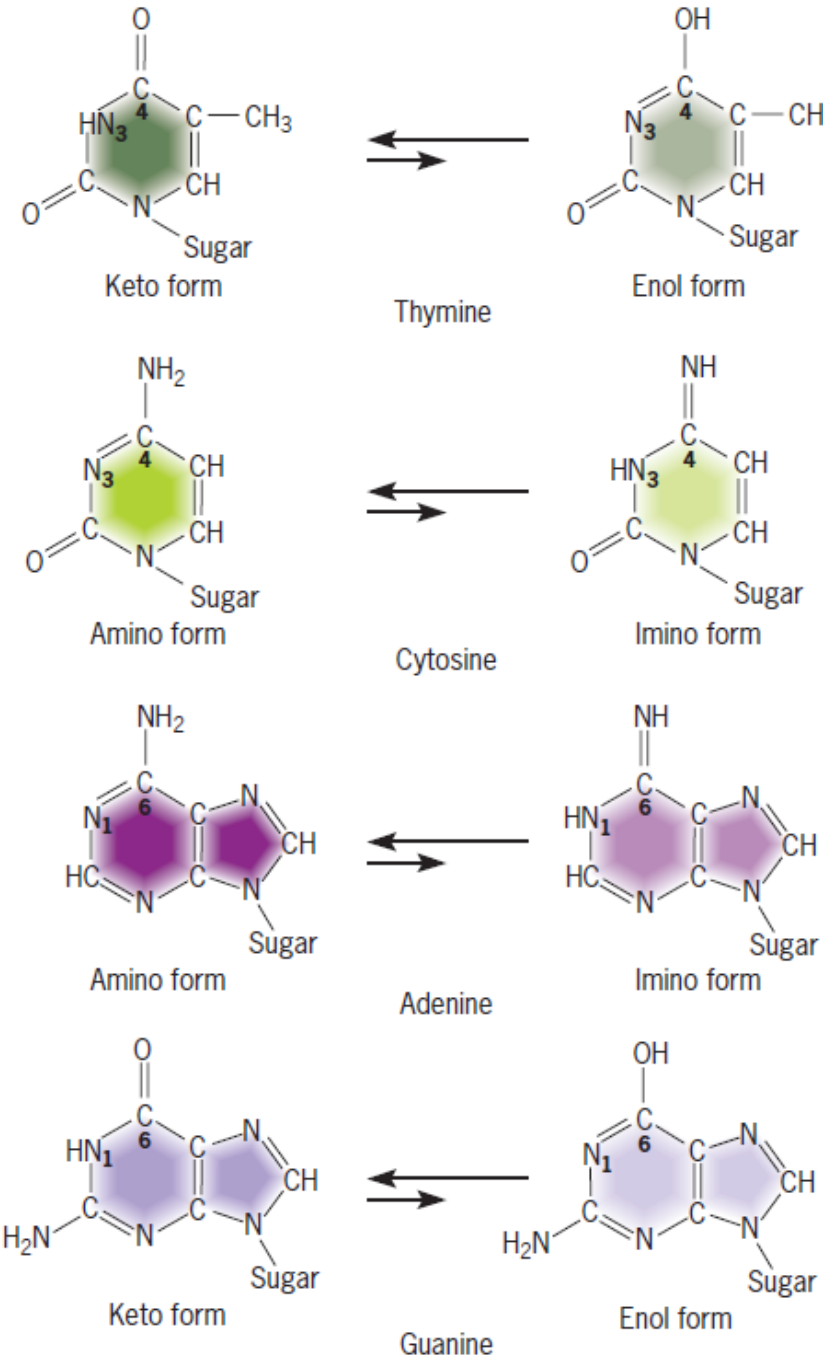
- **Spontaneous**
  - A spontaneous mutation is one that occurs as a result of natural processes in cells, for example DNA replication errors. These can be distinguished from induced mutations;
  - slipped strand mispairing can occur at homopolymeric runs (mono, di, or trinucleotide repeats)

Exposure to mutagens

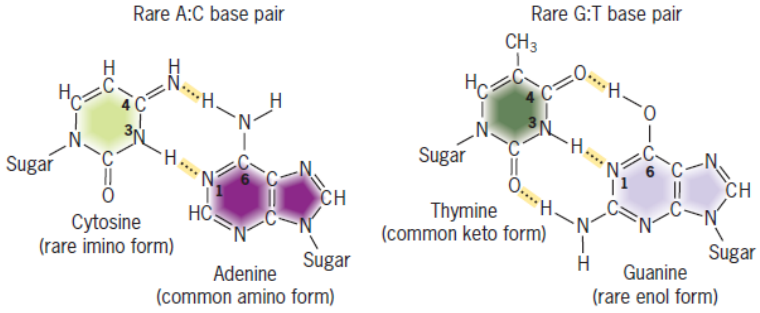
- Ionizing radiation—gamma rays and X-rays can cause single-strand or double-strand breaks in the sugar–phosphate backbone
- Ultraviolet radiation—(sunlight that can penetrate the ozone layer).
- Environmental chemicals—these include hydrocarbons (for example, in cigarette smoke/ barbecue). Attack by reactive oxygen species—highly reactive superoxide anions (O_2^-) and related molecules are generated as a by-product of oxidative metabolism in mitochondria.

•Spontaneous

- Exist in alternating forms called **tautomers**
  - Chemical modification of bases followed by mispairing
  - When the bases are present in their rare imino or enol states, they can form adenine–cytosine and guanine–thymine base pairs
- Those that occur as a result of interaction of DNA with an outside agent or mutagen that causes DNA damage.
- Moreover, some sites on chromosomes are “hotspots” where mutations arise at a higher frequency than other regions of the DNA.



Hydrogen-bonded A:C and G:T base pairs that form when cytosine and guanine are in their rare imino and enol tautomeric forms.



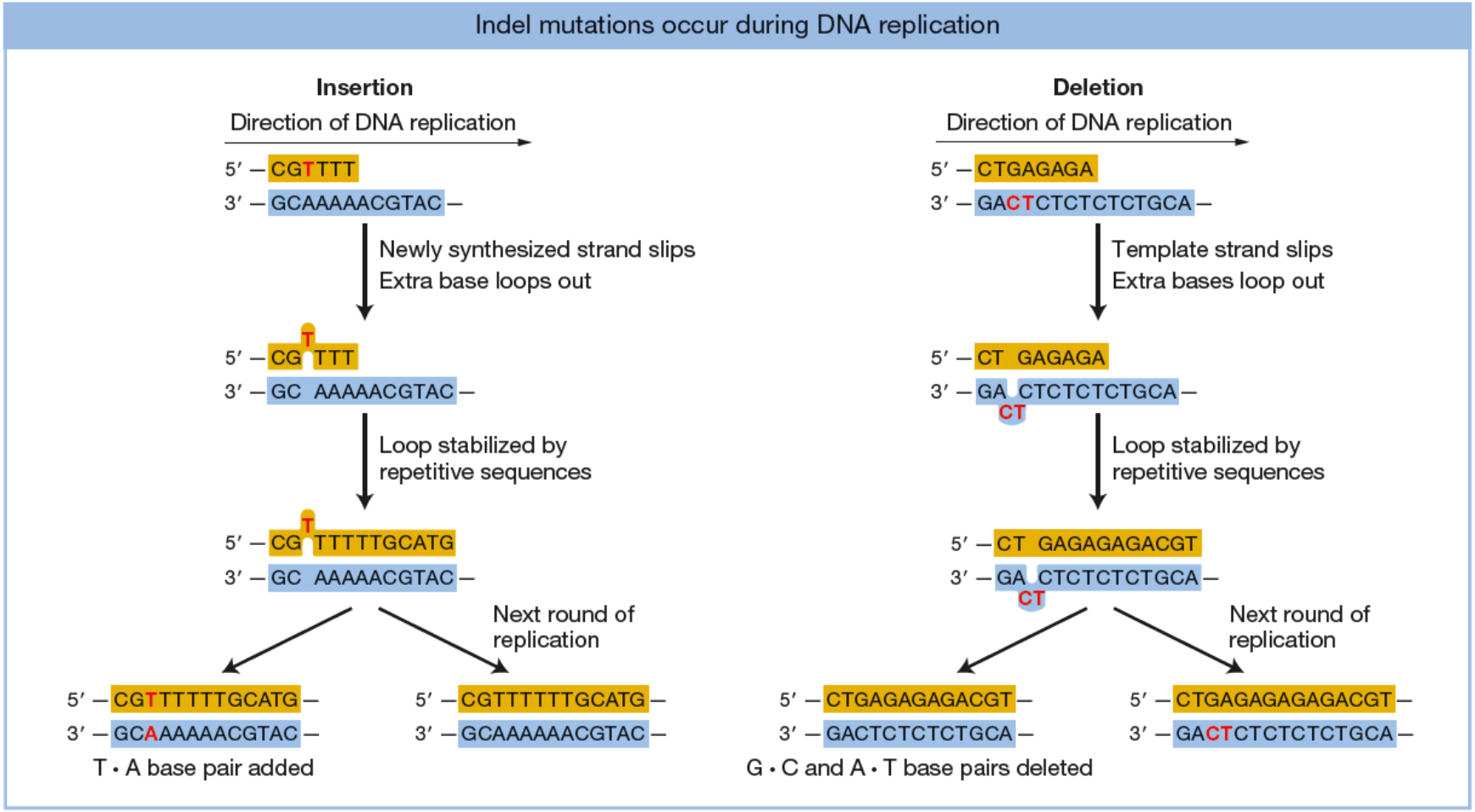


FIGURE 15-7 In the course of DNA replication, base insertions and deletions (indel mutations) are formed through the slipped mispairing of repeated sequences.

ANIMATED ART SaplingPlus

Molecular mechanism of mutation

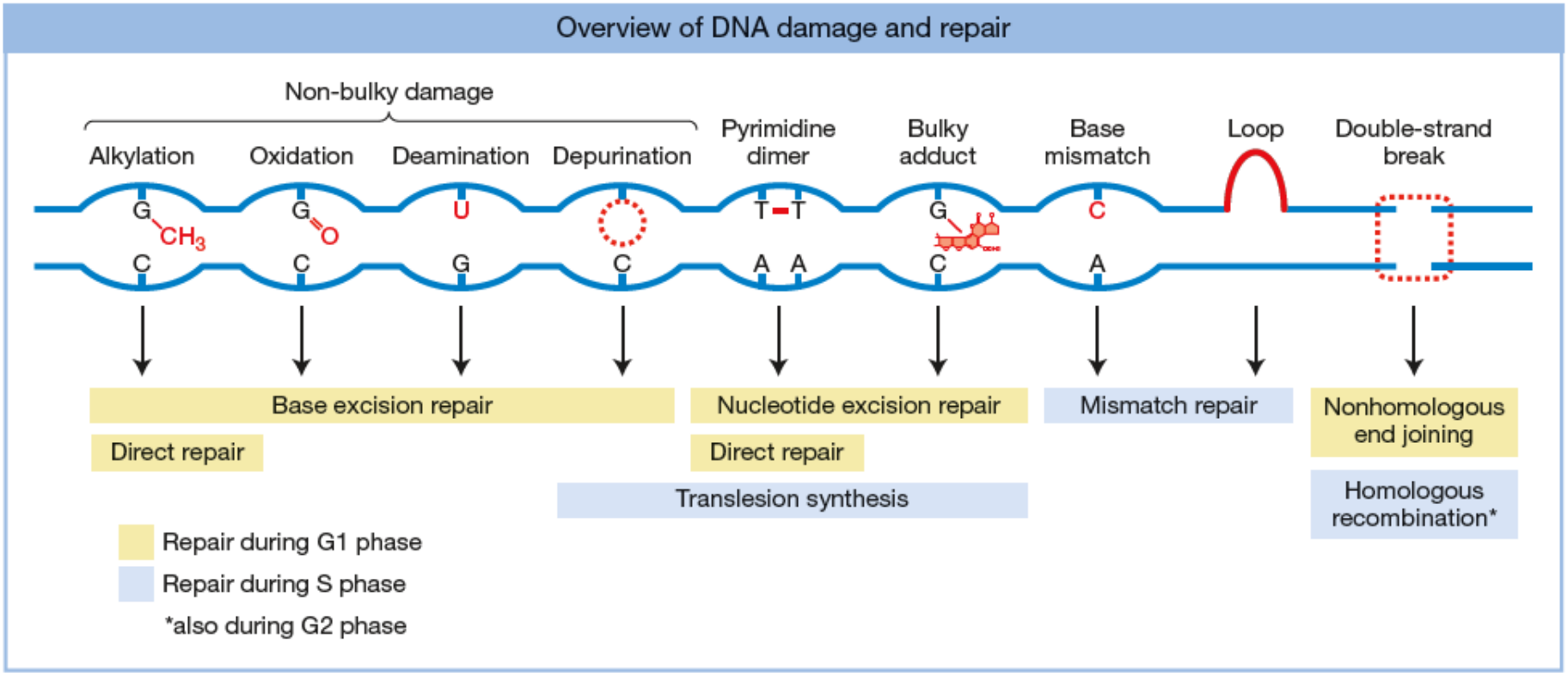


FIGURE 15-13 DNA repair mechanisms are paired with the types of DNA damage they act on. DNA damage is indicated in red. DNA repair mechanisms highlighted in tan function during G1 phase of the cell cycle, and those highlighted in blue function during S phase of the cell cycle. Homologous recombination also functions in G2 phase of the cell cycle.

- Even if the damage is recognized and excised, the repair machinery may not read the complementary strand accurately and, as a consequence, will create mutations by introducing incorrect bases.
- Thus, in contrast to replication-related DNA changes, which are usually corrected through proofreading mechanisms, nucleotide changes introduced by DNA damage and repair often result in permanent mutations.

TABLE 12.1

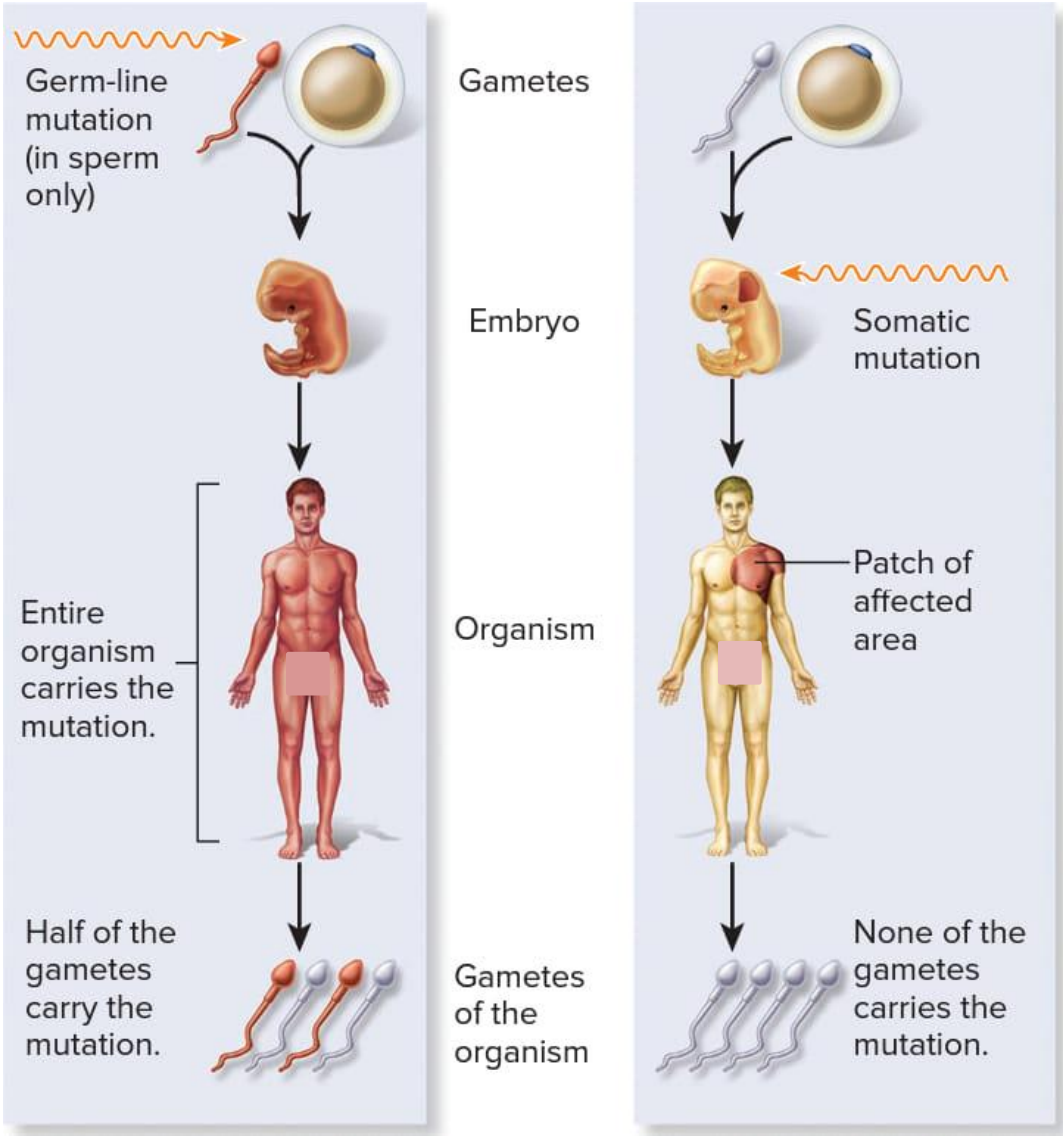
Major Types of Mutations and Their Distinguishing Features

Basis of classification	Major types of mutations	Major features
Origin	Spontaneous	Occurs in absence of known mutagen
	Induced	Occurs in presence of known mutagen
Cell type	Somatic	Occurs in nonreproductive cells
	Germ-line	Occurs in reproductive cells
Expression	Conditional	Expressed only under restrictive conditions (such as high temperature)
	Unconditional	Expressed under permissive conditions as well as restrictive conditions
Effect on function	Loss-of-function (knockout, null)	Eliminates normal function
	Hypomorphic (leaky)	Reduces normal function
	Hypermorphic	Increases normal function
	Gain-of-function (ectopic expression)	Expressed at incorrect time or in inappropriate cell types
Molecular change	Base substitution	One base pair in duplex DNA replaced with a different base pair
	Transition	Pyrimidine (T or C) to pyrimidine, or purine (A or G) to purine
	Transversion	Pyrimidine (T or C) to purine, or purine (A or G) to pyrimidine
	Insertion	One or more extra nucleotides present
	Deletion	One or more missing nucleotides
Effect on translation	Synonymous (silent)	No change in amino acid encoded
	Missense (nonsynonymous)	Change in amino acid encoded
	Nonsense (termination)	Creates translational termination codon (UAA, UAG, or UGA)
	Frameshift	Shifts triplet reading of codons out of correct phase

Table 12.01: Major types of mutations and their distinguishing features.

Two categories of mutations

- A mutation is change in a DNA sequence is rare in a population and typically affects the phenotype
- Somatic Mutations
  - Occur in cells of the body that do not form gametes
  - Occurs in mitosis: can lead to mosaicism
  - Is not transmitted to future generations
  - Tumor – uncontrolled growth
- Germ-line Mutations
  - Occur in cells that produce gametes
  - Occurs during meiosis
  - Transmitted to future generations - inherited



Types of mutations and their phenotypic consequences

Mutations are of fundamental importance in molecular biology for several reasons:

1 As noted above, mutations are important as the major source of genetic variation.

2 Mutations may have deleterious or (rarely) advantageous consequences to an organism or its descendants. Mutations in germ cells can lead to heritable genetic disorders, while mutations in somatic cells may lead to acquired diseases such as cancer or neurodegenerative disorders.

3 Mutant organisms are important tools for molecular biologists in characterizing the genes involved in cellular processes.

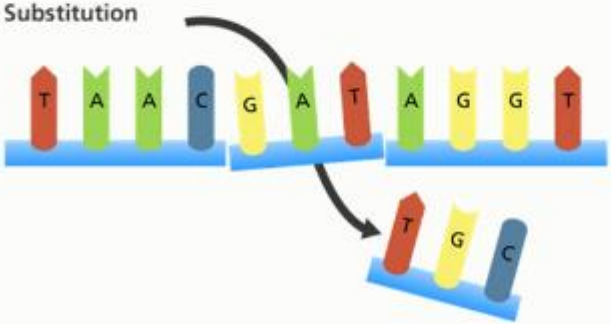
Small- scale mutations/ variations

- The simplest type of mutation is a nucleotide substitution.
- Mutations that alter a single nucleotide are called point mutations.

• **Substitution** – when one or more bases in the sequence is replaced by the same number of bases (for example, a cytosine substituted for an adenine).

Original sequence

Substitution

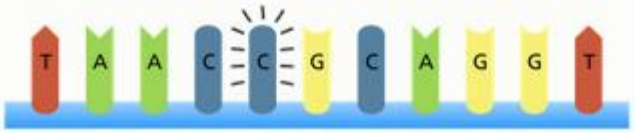


Small-scale mutations

• **Point mutation** – a change in one base in the DNA sequence.

Original sequence

Point mutation



Types of mutations

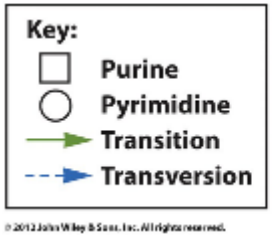
At the Nucleotide Level (Point mutations):

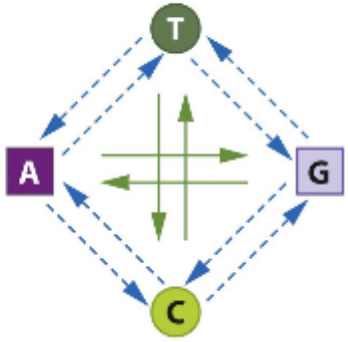
- A base-pair substitution replaces one nucleotide and its partner with another pair of nucleotides is called base-pair substitution mutations.
- Two forms:
  - Transition: replaces a pyrimidine with another pyrimidine or a purine for another purine
  - Transversion: replaces a pyrimidine with a purine or or vice verse

Table 7.1 Types of nucleotide substitutions.

Nucleotide substitution	Mutation
Transition mutation Pyrimidine → pyrimidine	T → C or C → T
Purine → purine	A → G or G → A
Transversion mutation Pyrimidine → purine	T → A, T → G, C → A, or C → G
Purine → pyrimidine	A → T, A → C, G → T, or G → C

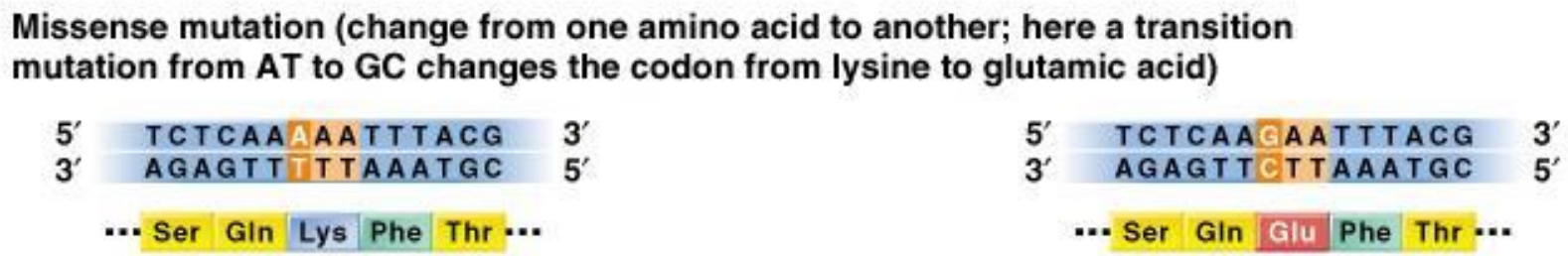
Twelve different base substitutions can occur in DNA.





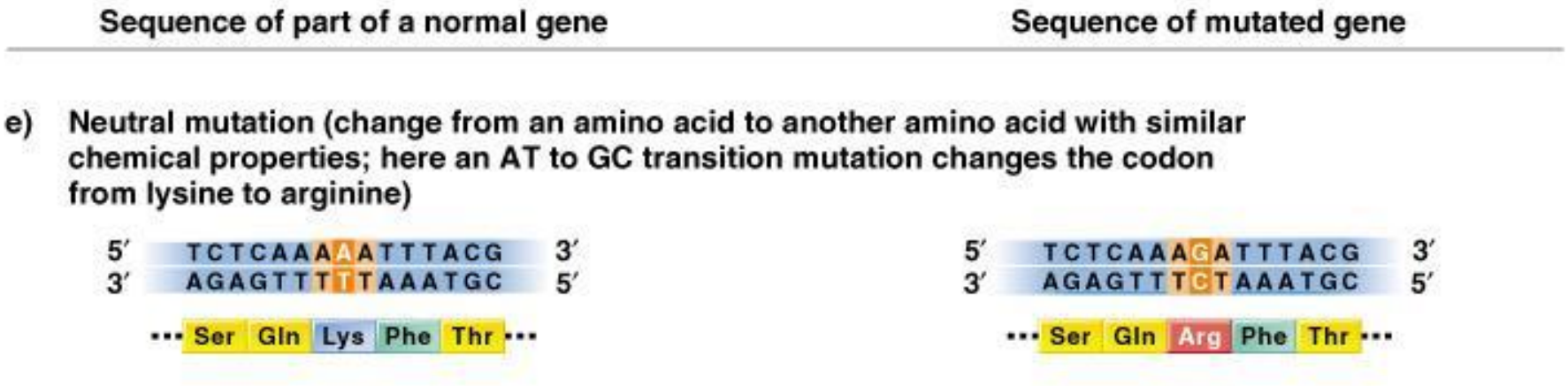
Missense mutations

- Nucleotide substitutions in protein-coding regions that do result in changed amino acids are called nonsynonymous mutations or missense mutations.
- May alter the biological properties of the protein.
- A classic example of a phenotypic effect of a single amino acid change is phenylketonuria, a disorder that can be caused by a base substitution in the phenylalanine hydrolase gene
- The mutation that leads to PKU is a transversion from a G to a C at codon 413
- This mutation is also a missense mutation in that it will result in an amino acid substitution in the corresponding protein from Pro413 to Arg 413
- This mutation will result in a PAH enzyme that is no longer functional, and will not be able to metabolize phenylalanine
- This leads to a variety of developmental defects including severe mental retardation

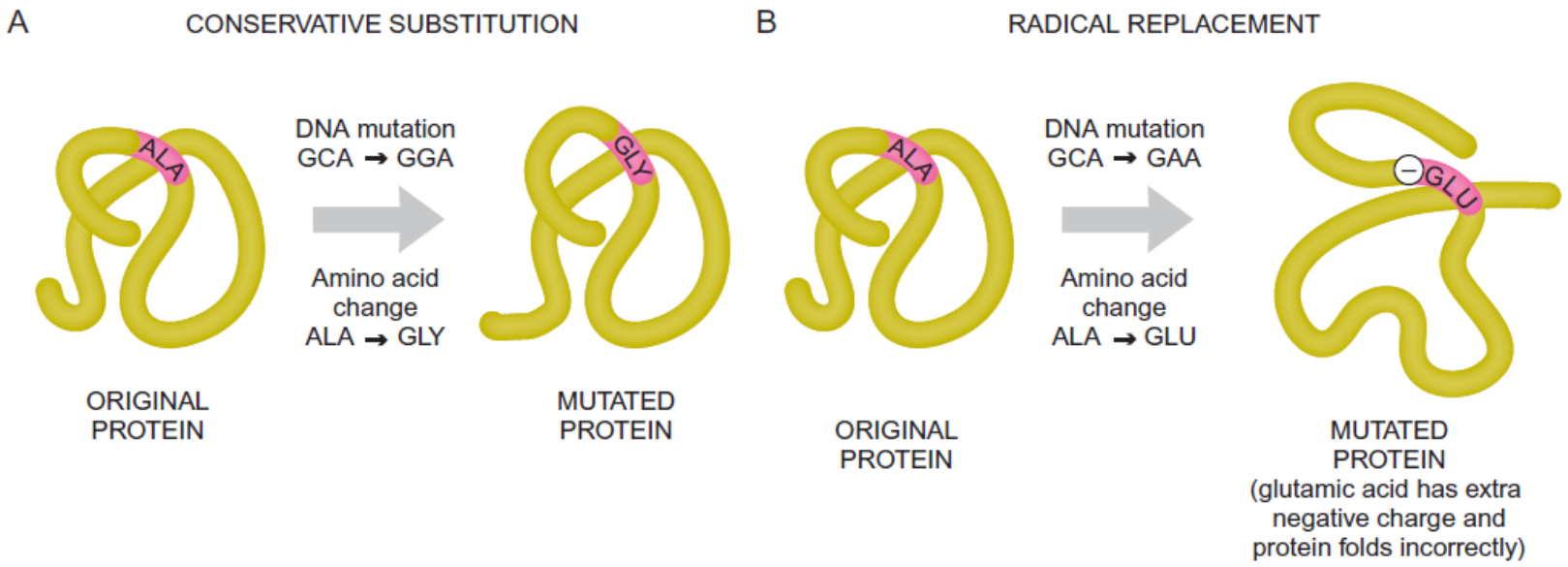


Missense mutations

- Neutral nonsynonymous mutation: Base pair substitution results in substitution of an amino acid with similar chemical properties (protein function is not altered).



  - Conservative – chemical properties of mutant amino acid are similar to the original amino acid
    - e.g. aspartic acid [(-)charged] → glutamic acid [(-)charged]
  - Nonconservative – chemical properties of mutant amino acid are different from original amino acid
    - e.g. aspartic acid [(-)charged] → alanine (uncharged)


A
CONSERVATIVE SUBSTITUTION
ALA
DNA mutation
GCA → GGA
Amino acid
change
ALA → GLY
GLY
ORIGINAL
PROTEIN
MUTATED
PROTEIN
B
RADICAL REPLACEMENT
ALA
DNA mutation
GCA → GAA
Amino acid
change
ALA → GLU
GLU
ORIGINAL
PROTEIN
MUTATED
PROTEIN
(glutamic acid has extra
negative charge and
protein folds incorrectly)

Silent mutations

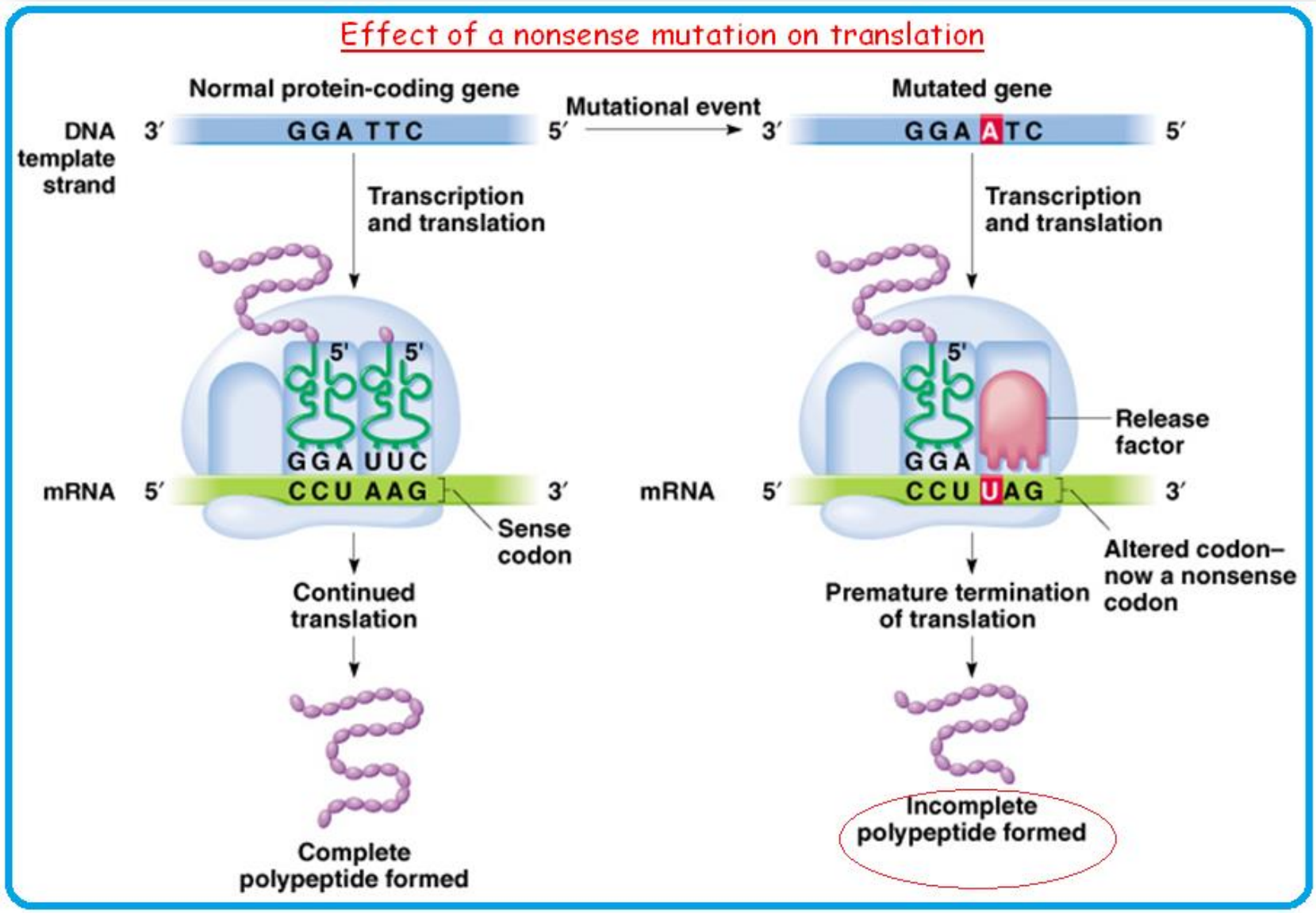
- Nucleotide substitutions in a protein-coding gene may or may not change the amino acid in the encoded protein.
- Mutations that change the nucleotide sequence without changing the amino acid sequence are called synonymous mutations or silent mutations.
- Mutational changes in nucleotides that are outside of coding regions can also be silent. However, some noncoding sequences do have essential functions in gene regulation and, in this case, mutations in these sequences would have phenotypic effects.

Silent mutation (change in codon such that the same amino acid is specified; here an AT-to-GC transition in the third position of the codon gives a codon that still encodes lysine)

Nonsense mutations

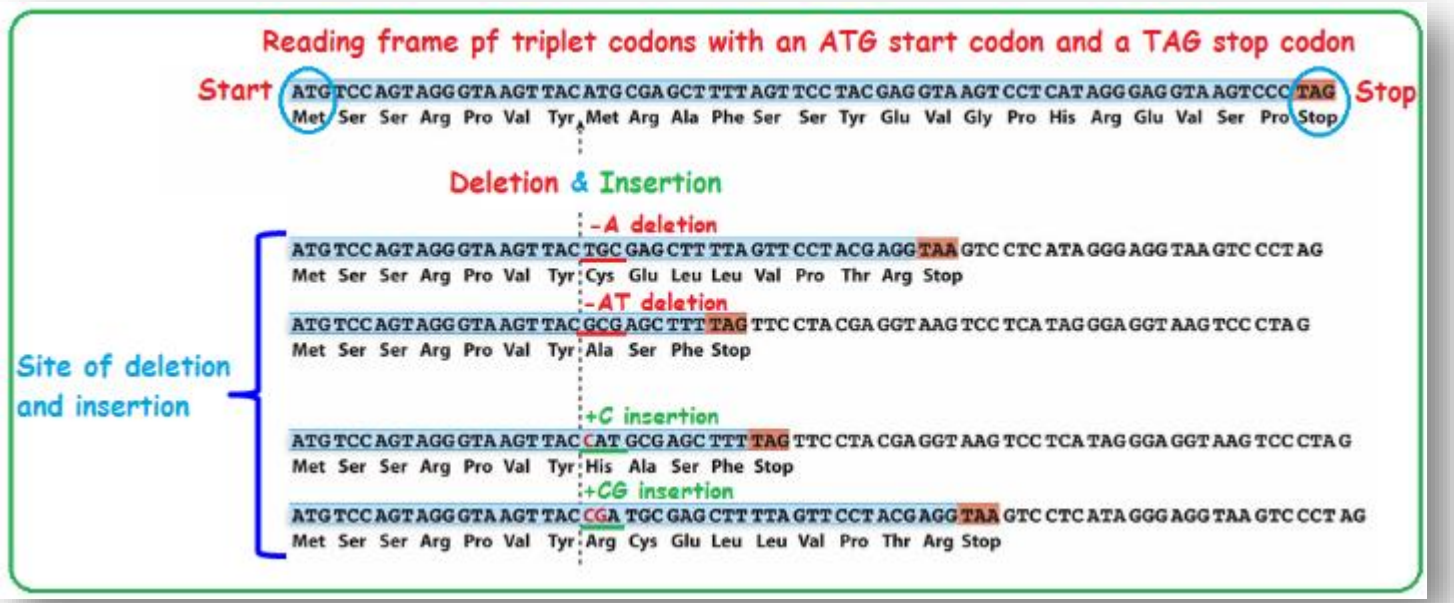
- Nonsense mutations change codon that encodes an amino acid to a stop codon (UGA, UAG, or UAA)
- Nonsense mutations (genes containing premature termination codons often cause production of the truncated protein that might be predicted.
- Cells have a mechanism, nonsense-mediated decay (NMD), that detects mRNAs containing premature termination codons and degrades them. Thus, the usual result of a nonsense mutation is to prevent any expression of the gene
- Nearly always a nonfunctional product.

d) Nonsense mutation (change from an amino acid to a stop codon; here a transversion mutation from AT to TA changes the codon from lysine to UAA stop codon)

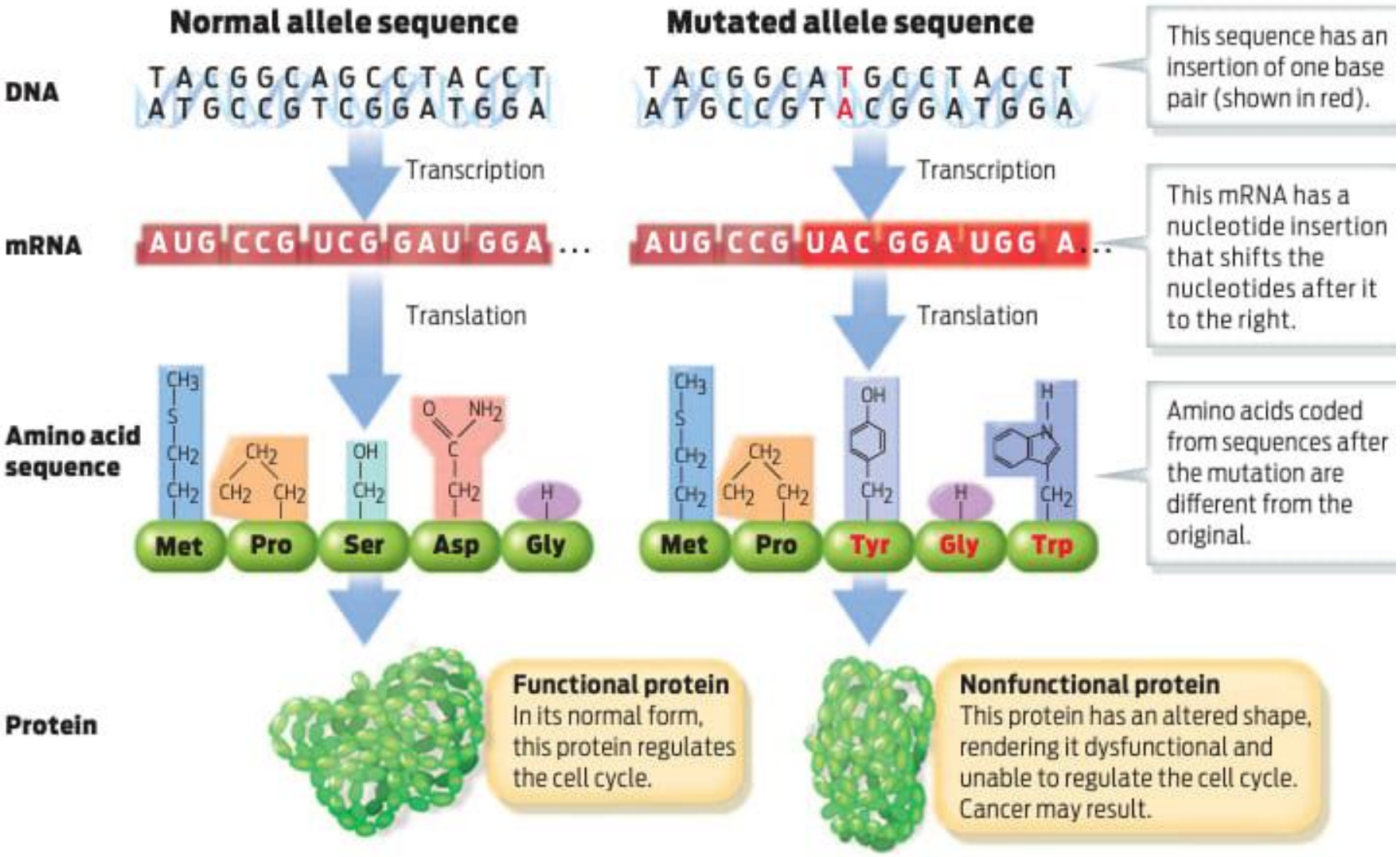

Effect of a nonsense mutation on translation
Normal protein-coding gene
Mutational event
Mutated gene
DNA template strand
3′
GGA TTC
5′
3′
GGA ATC
5′
Transcription and translation
Transcription and translation
5'
5'
5'
GGA UUC
GGA
mRNA
5′
CCU AAG
3′
mRNA
5′
CCU UAG
3′
Release factor
Sense codon
Altered codon–now a nonsense codon
Continued translation
Premature termination of translation
Complete polypeptide formed
Incomplete polypeptide formed

Insertions or deletions can cause frameshift mutations

- Insertions or deletions of nucleotides can also occur in DNA, but at a rate considerably lower than that of nucleotide substitution.
- Insertion and deletion mutations are collectively referred to as indels.
- The DNA sequence from the start codon to the stop codon is referred to as a reading frame.
- Because nucleotides are decoded in triplets, an indel mutation of only one or two base pairs in the coding sequence of a protein throws off the reading frame after the mutation, resulting in a frameshift mutation.



- A classic example of a phenotypic effect of a **small deletion is the change responsible for the human hereditary disease cystic fibrosis**.
- The deletion of three base pairs in the nucleotide sequence of **the cystic fibrosis transmembrane conductance regulator (CFTR)** gene results in **the loss of the codon for phenylalanine .**
- If the length of an **insertion or deletion is not an exact multiple of three nucleotides, the mutation shifts the phase in which the ribosome reads the triplet codons and, consequently, alters all of the amino acids downstream from the site of the mutation.**
- Such **mutations are called frameshift mutations** because they "shift" the reading frame of the codons in the mRNA.
- Usually leads to production of a nonfunctional protein.


Normal allele sequence
Mutated allele sequence
DNA
T A C G G C A G C C T A C C T
A T G C C G T C G G A T G G A
T A C G G C A T G C C T A C C T
A T G C C G T A C G G A T G G A
This sequence has an insertion of one base pair (shown in red).
Transcription
Transcription
mRNA
AUG CCG UCG GAU GGA ...
AUG CCG UAC GGA UGG A...
This mRNA has a nucleotide insertion that shifts the nucleotides after it to the right.
Translation
Translation
Amino acid sequence
Met
Pro
Ser
Asp
Gly
Met
Pro
Tyr
Gly
Trp
Amino acids coded from sequences after the mutation are different from the original.
Protein
Functional protein
In its normal form, this protein regulates the cell cycle.
Nonfunctional protein
This protein has an altered shape, rendering it dysfunctional and unable to regulate the cell cycle. Cancer may result.

Regulatory Mutations

Occur in noncoding regions of genes

Three types are commonly recognized: promoter mutations, Polyadenylation, splicing mutations, and cryptic splice sites.

Splice Site Mutations

Two general classes of splicing mutation have been described.

- A mutation can also alter splice junctions in eukaryotes (splicing mutations)
    - Leads to splicing abnormalities by changing the donor and acceptor site of splicing sites-> interfere with normal RNA splicing at that site.
- A second class of splicing mutations involves intron base substitutions (cryptic splice sites), such mutations create alternative donor or acceptor sites that compete with the normal sites during RNA processing.
    - Cryptic donor or acceptor are new sites of splicing donors or acceptors.

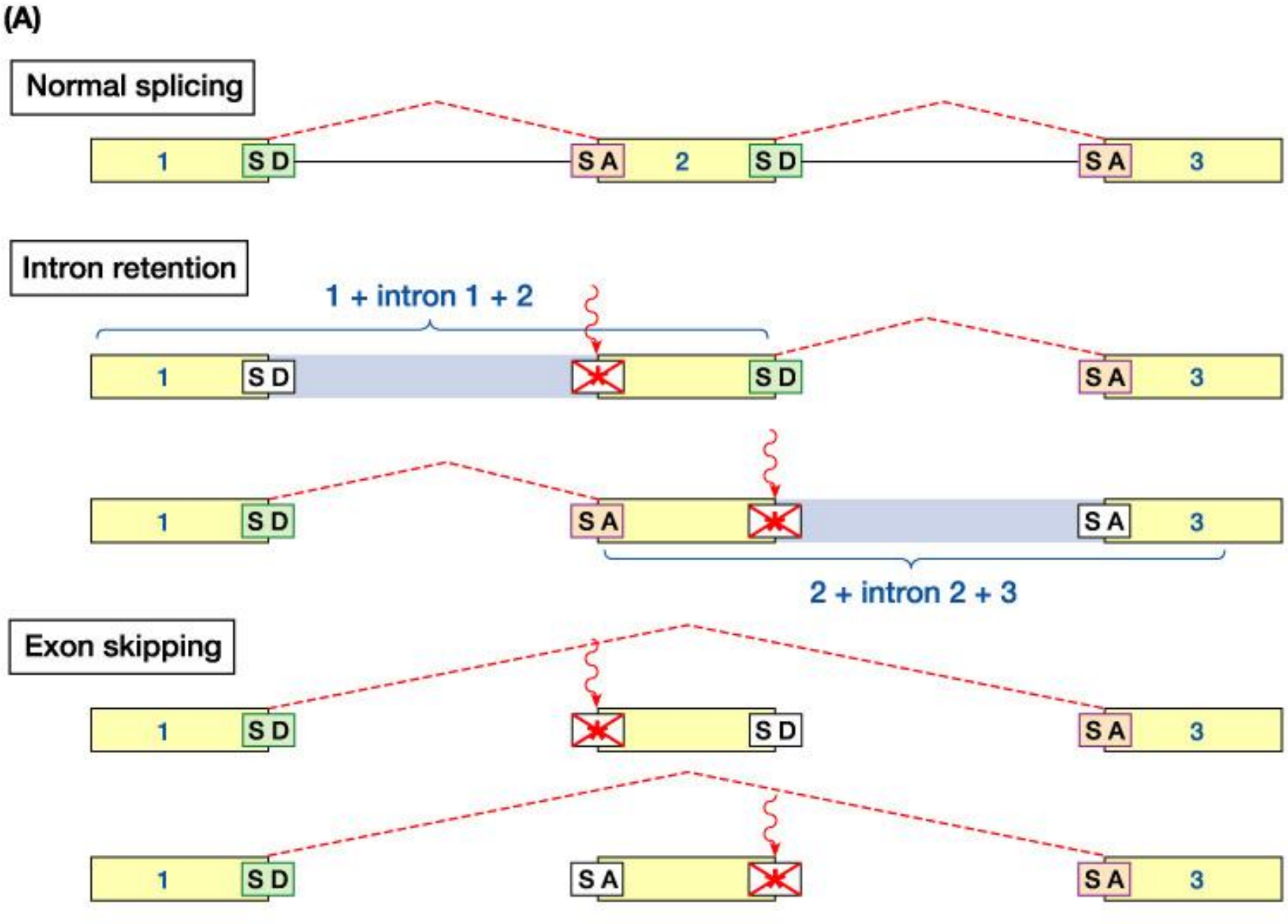


Figure 11-11 part 1 of 2 Human Molecular Genetics, 3/e. (© Garland Science 2004)

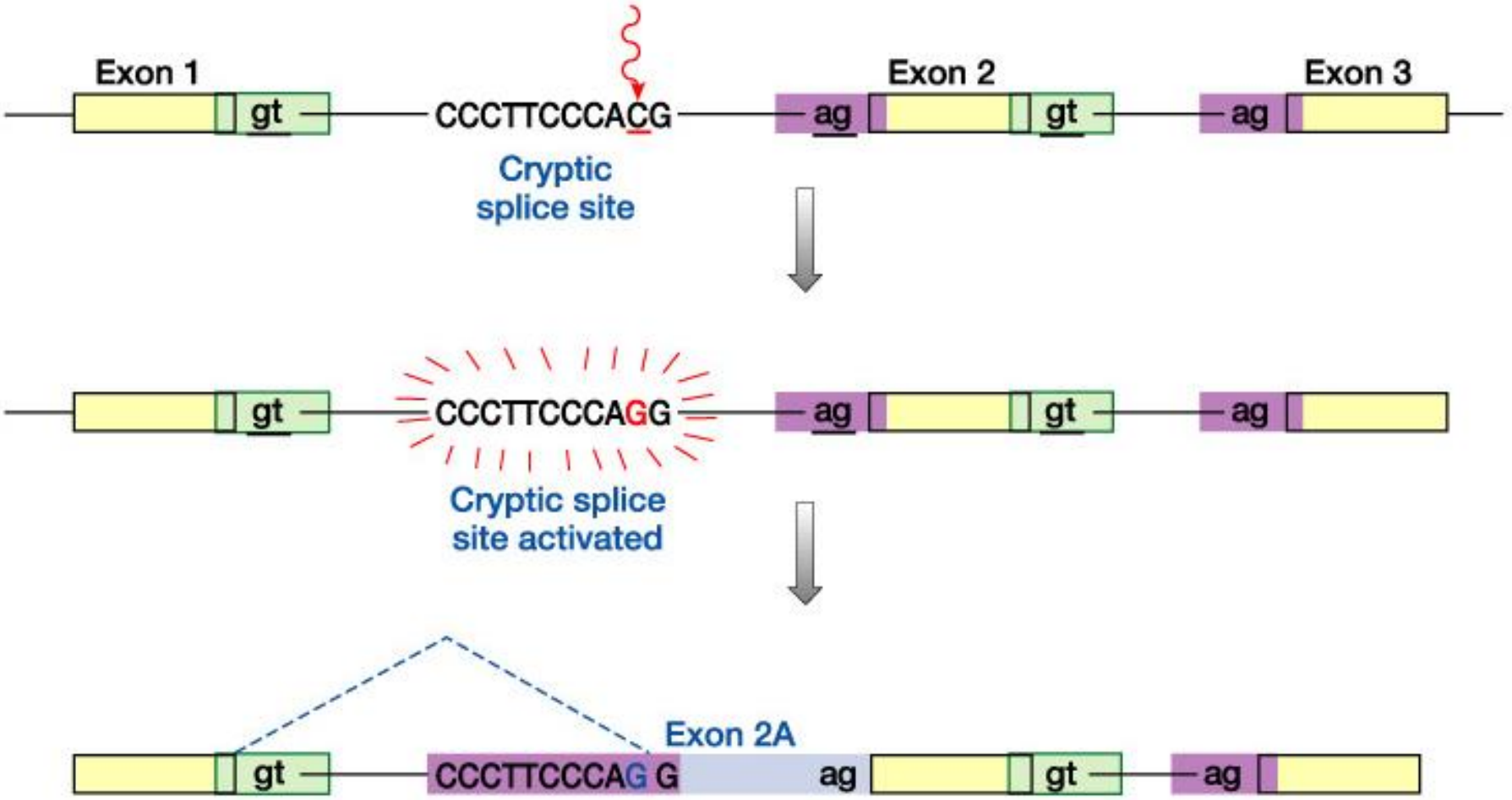


Figure 11-13 Human Molecular Genetics, 3/e. (© Garland Science 2004)

33

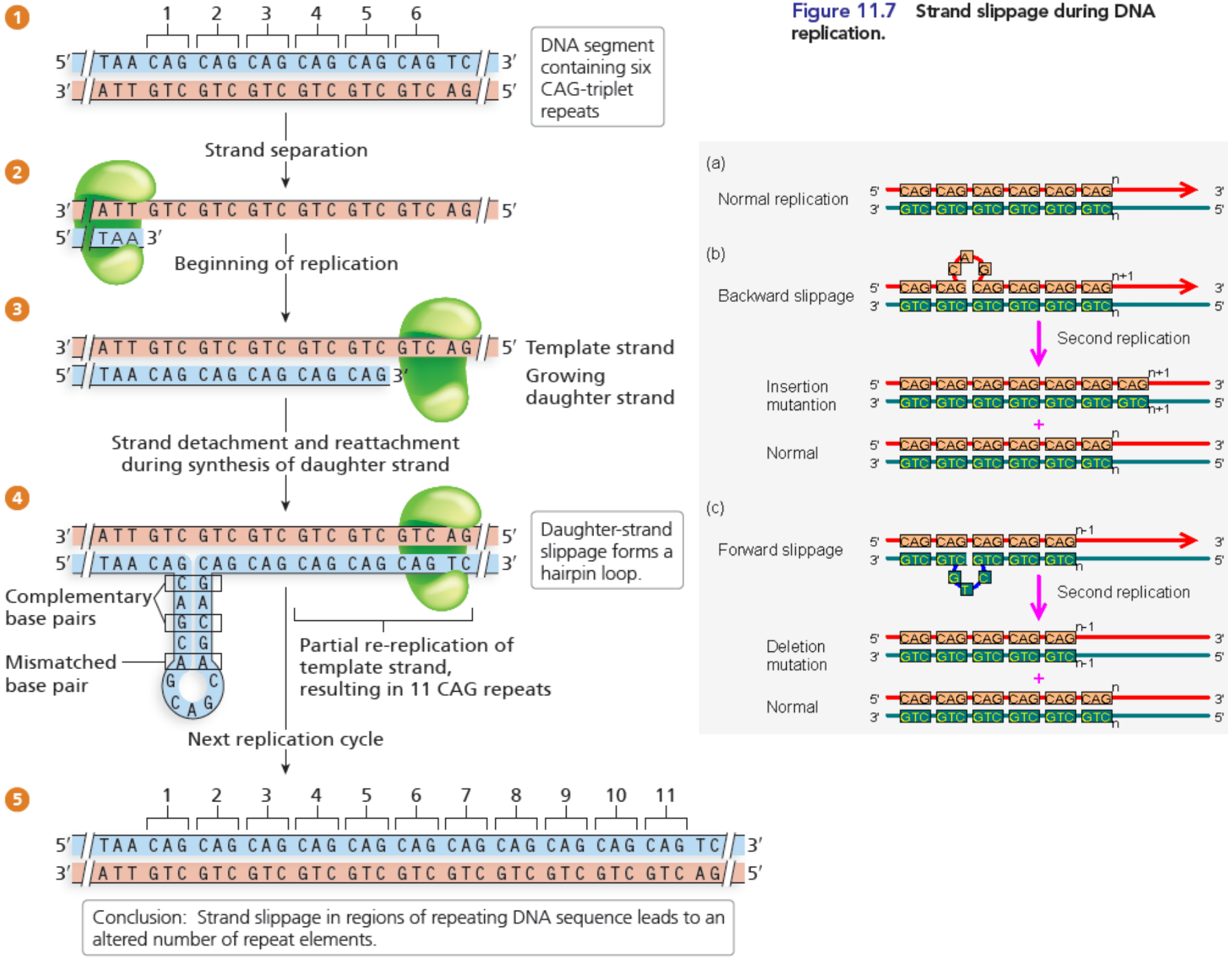


Figure 11.7 **Strand slippage during DNA replication.**

Gene and Variant nomenclature

Genes: https://www.genenames.org/
Variant: https://varnomen.hgvs.org/

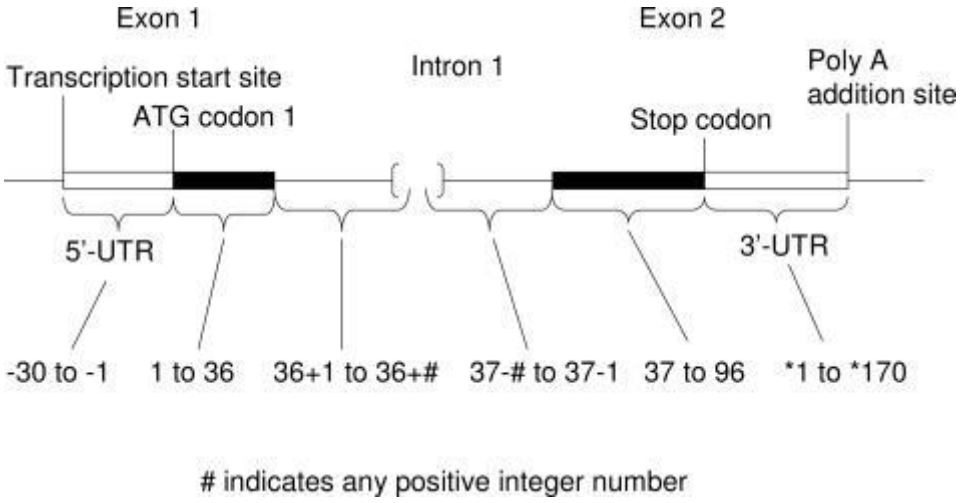
- The HGVS nomenclature guidelines are used worldwide for genetic variant interpretation
- The Human Genome Variation Society (HGVS) nomenclature standard was developed to prevent the misinterpretation of variants in DNA, RNA, and protein sequences. The HGVS nomenclature standard is used worldwide, especially in clinical diagnostics, and is authorized by the Human Genome Organisation (HUGO).

Reference Sequence Types

- Depending on the variants to be reported, different reference sequence files are used at the DNA, RNA or protein level. It is mandatory to indicate the type of reference sequence file using a**prefix**preceding the variant description. Approved reference sequence types are**c.**,**g.**,**m.**,**n.**,**o.**,**p.**and**r.**:
- •DNA
- •**g.**=linear genomic reference sequence
- •**o.**=circular genomic reference sequence
- •**m.**=mitochondrial reference(special case of a circular genomic reference sequence)
- •**c.**=coding DNA reference sequence(based on a protein coding transcript)
- •**n.**=non-coding DNA reference sequence(based on a transcript not coding for a protein)

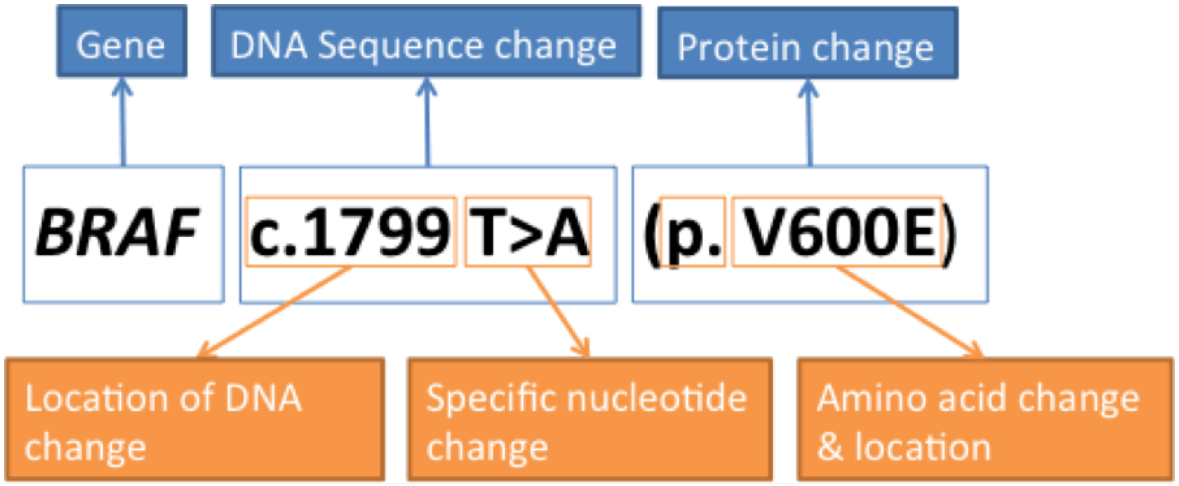
Variant nomenclature: cDNA

- Nucleotide 1 is the A of the ATG initiation codon (there is no c.0)
- The nucleotide 5’ of the ATG initiation codon is -1, the previous -2, etc.
- The nucleotide 3’ of the stop codon is *1, the next *2, etc.
- Intronic nucleotides
- 5’ end of the intron: the number of last coding nucleotide of the preceding exon, a plus sign and the position within in the intron, e.g., c.36+1G, c.36+2T
- 3’ end of the intron: the number of the first coding nucleotide of the following exon, a minus sign and the position upstream in the intron, e.g. c.37-1G, c.37-2A



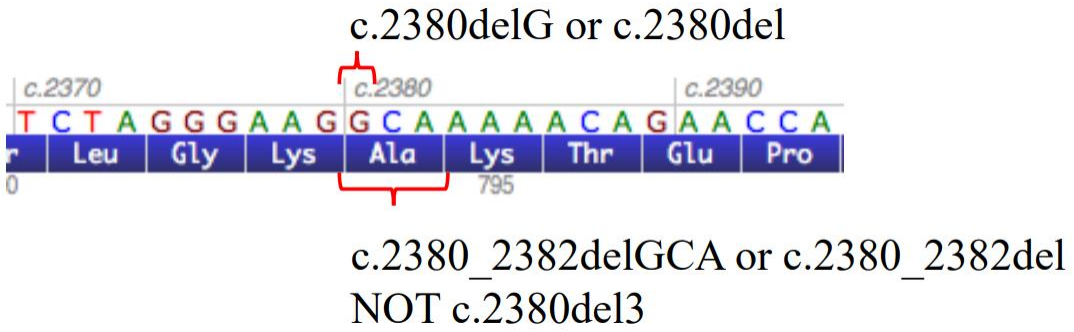
Symbols and abbreviations

>	c.4375C>T	**Substitution** of the C nucleotide at position c.4375 with a T
del	c.4375_4379del or c.4375_4379delCGATT	Nucleotides from position c.4375 to c.4379 **deleted**
dup	c.4375_4385dup or c.4375_4385dupCGATTATTCCA	Nucleotides from position c.4375 to c.4385 **duplicated**
ins	c.4375_4376insACCT	ACCT **inserted** between positions c.4375 and c.4376
delins	c.4375_4376delinsACTT or c.4375_4376delCGinsAGTT	Nucleotides from position c.4375 to c.4376 (CG) are **deleted** and **replaced** by ACTT



"c." prefix denotes standard variant nomenclature based on coding DNA reference sequences
"p." prefix denotes standard variant nomenclature based on protein-level amino acid sequences

- Format: **"prefix""position(s)_deleted""del",** e.g. g.123_127del



- For all descriptions the **most 3' position** possible of the reference sequence is arbitrarily assigned to have been changed (3'rule)

Variant nomenclature: protein

- 3-letter amino acid code is preferred to describe the amino acid residues (Lys vs. K for lysine)
- For all descriptions the most C-terminal position possible is arbitrarily assigned to have been changed
- Methionine encoded by the translation initiation site (start codon) is numbered as residue 1 ("Met1" or " M1")
- "Ter" or "*" designating a translation termination codon (some labs use X)

Protein nomenclature

- **Silent changes:** p.Leu54Leu or p.=
- **Substitutions:** p.Trp26Cys • Nonsense variant: p.Trp26Ter or p.Trp26*
- **No-stop change:** p.Ter110GlnextTer17 or p.*110Glnext*17
- **In-frame deletions:** p.Gln8del or p.Cys28_Met30del
- **Duplications:** p.Gly4_Gln6dup
- **Insertions:** p.Lys2_Met3insGlnSerLys
- **Frameshifts:** short description: p.Arg97fs

  long description: p.Arg97Profs*23

where the “Arg97Pro” describes the substitution of Arg for Pro at position 97, “fs” indicating the frameshift and the “*23” describes the position of the translational termination (stop) codon in the new reading frame (starting with proline as amino acid #1)