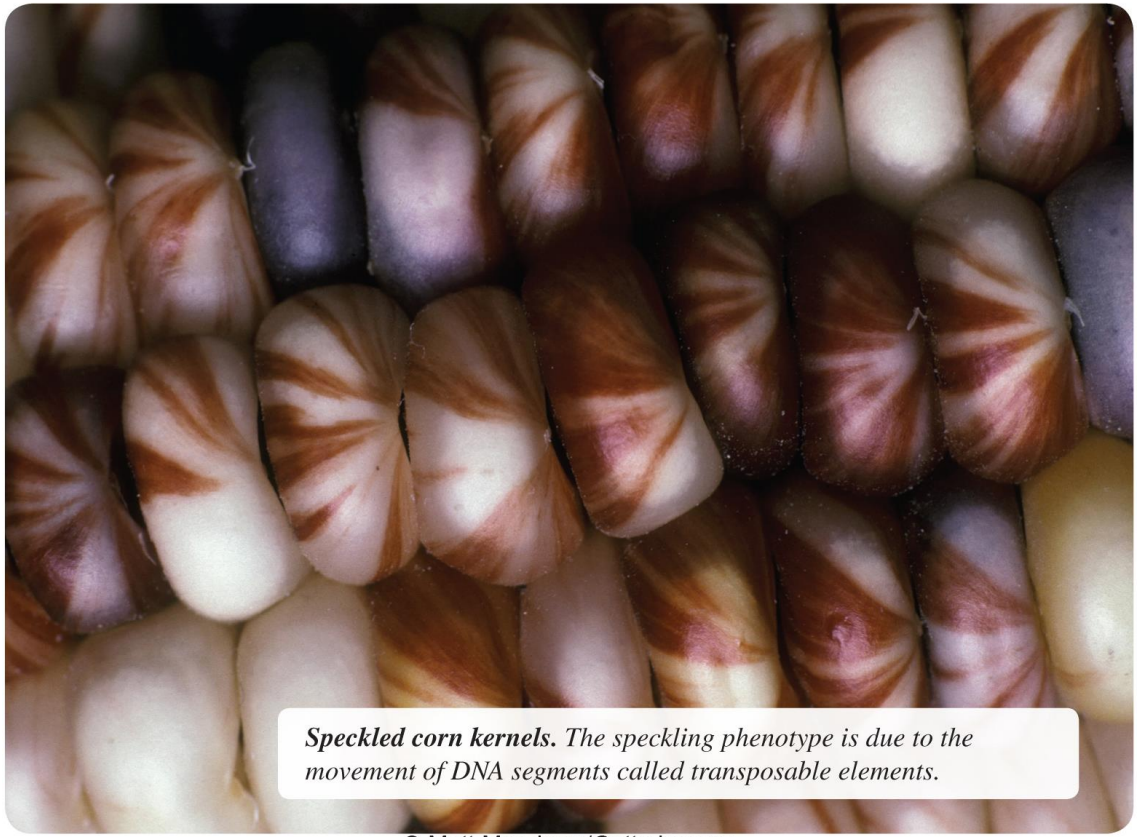
Chapter 19

Recombination, Immunogenetics, and Transposition



Speckled corn kernels. *The speckling phenotype is due to the movement of DNA segments called transposable elements.*

The speckling phenotype in these corn kernels is due to the movement of DNA segments called transposable elements.

- This chapter looks at a three molecular processes in which segments of chromosomal DNA become rearranged

1. **Homologous recombination** – Similar DNA segments break and rejoin to form new combinations during crossing-over in meiosis

2. **Site-specific recombination** – Certain genes of the immune system recombine nonhomologous DNA segments at specific sites, to generate antibody diversity

3. **Transposition** – Small segments of DNA called transposable elements move themselves to new locations in the chromosomes

19.1 Homologous Recombination

- The Holiday model and the double-strand break model for homologous recombination
- How gene conversion can occur via mismatch repair and DNA gap repair synthesis

Homologous Recombination

- Homologous recombination is found in all species
  - The cells of any given species may have more than one molecular mechanism for homologous recombination

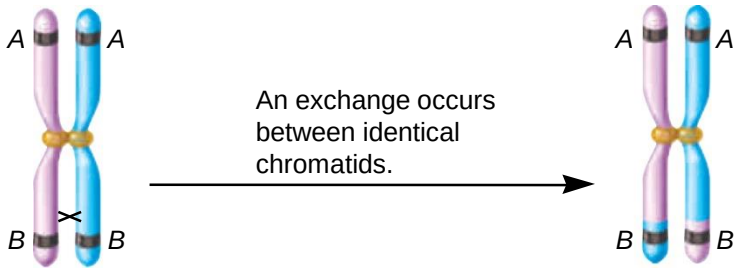
- The enzymology of homologous recombination is best understood in *E. coli*
  - Although haploid, they may have more than one copy of the chromosome, as well as during replication (good for DNA repair)

- In eukaryotes, occurs most frequently in meiosis but can occur in mitosis

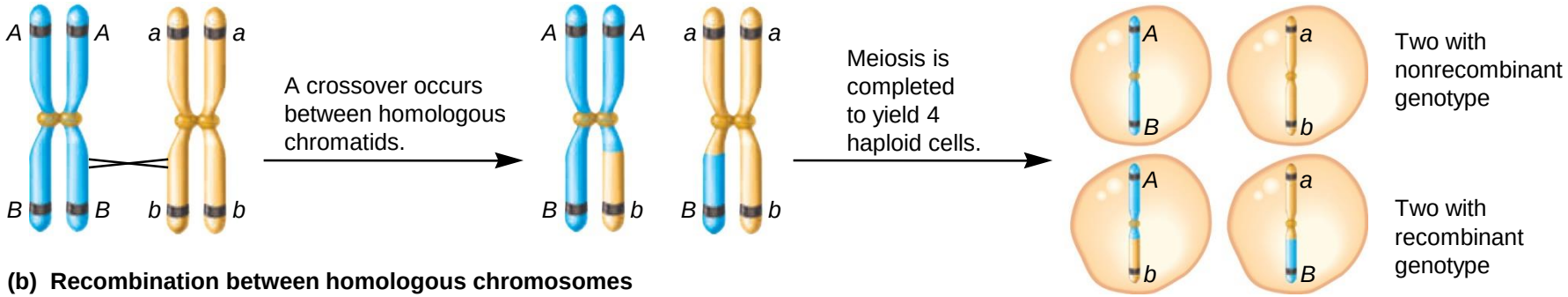
- Crossing over occurs in meiosis I
  - Exchange of DNA between non-sister chromatids of homologous chromosomes
  - Results in **genetic recombination**

- Crossing over that occurs between sister chromatids is called **sister chromatid exchange (SCE)**
  - Sister chromatids are genetically identical to each other
    - Therefore, SCE does not produce a new combination of alleles

    - Refer to Figure 19.1



(a) Sister chromatid exchange



(b) Recombination between homologous chromosomes

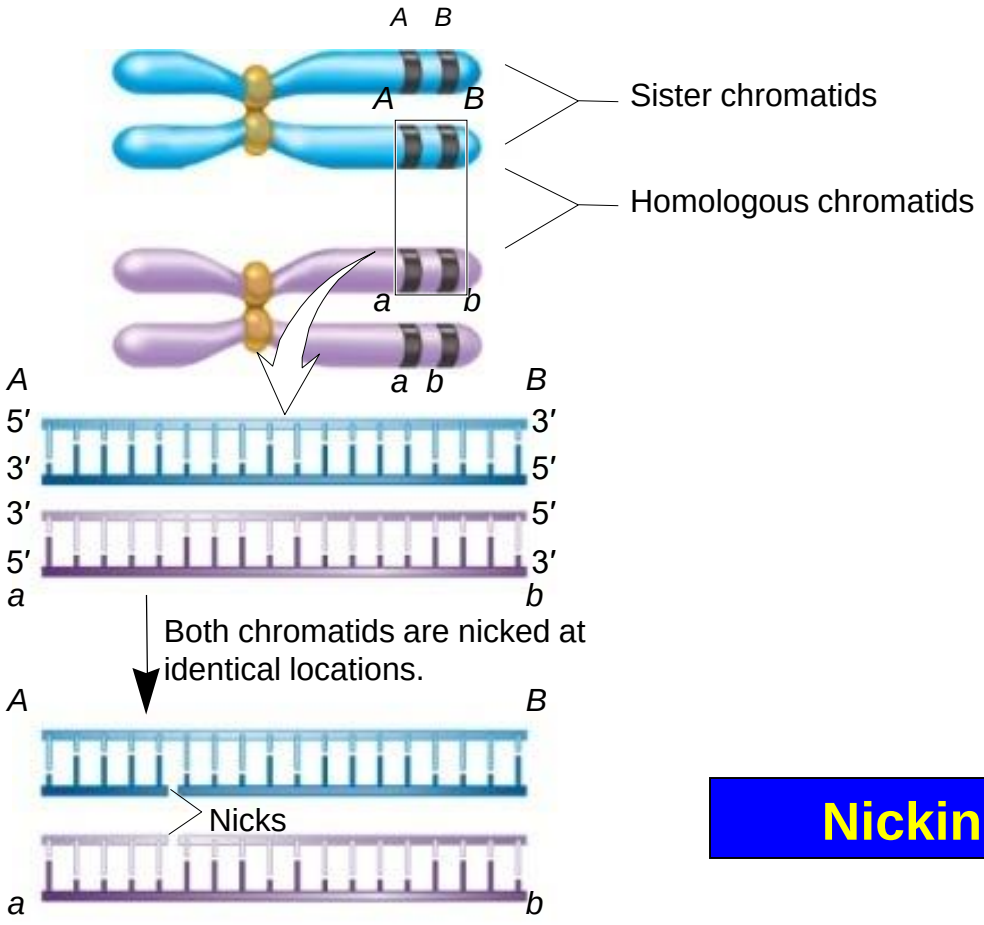
Figure 19.1

The Holliday Model for Homologous Recombination

- The first model of homologous recombination was deduced from genetic crosses in the fungus *Neurospora* by H. Zickler

- **Gene conversion** – Described unusual outcomes from the fungal crosses
  - Results not explained by mutations, too high a rate of occurrence – one allele was converted to the other allele

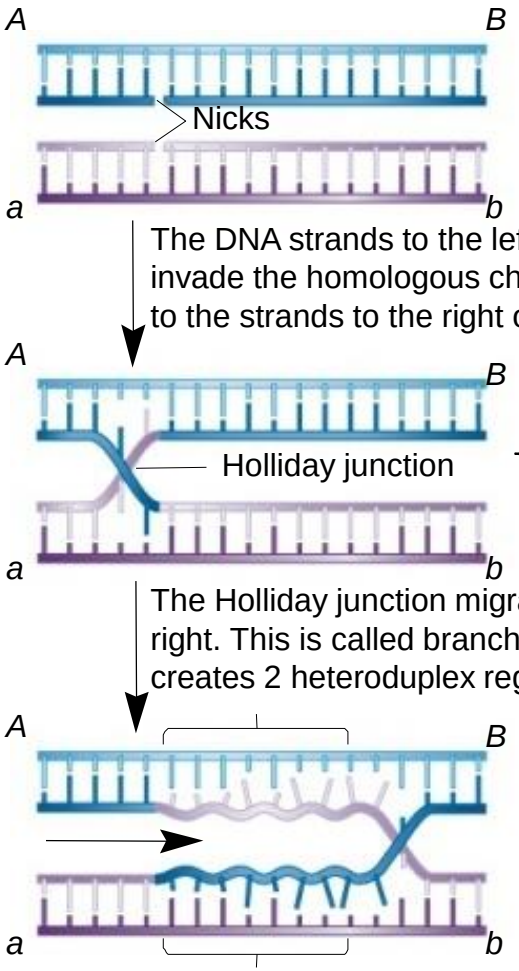
- Based on studies of gene conversion, Robin Holliday in 1964 proposed a model for homologous recombination
  - This model is shown in Figure 19.2



Alignment of homologous chromosomes

Nicking

Figure 19.2



Strand invasion

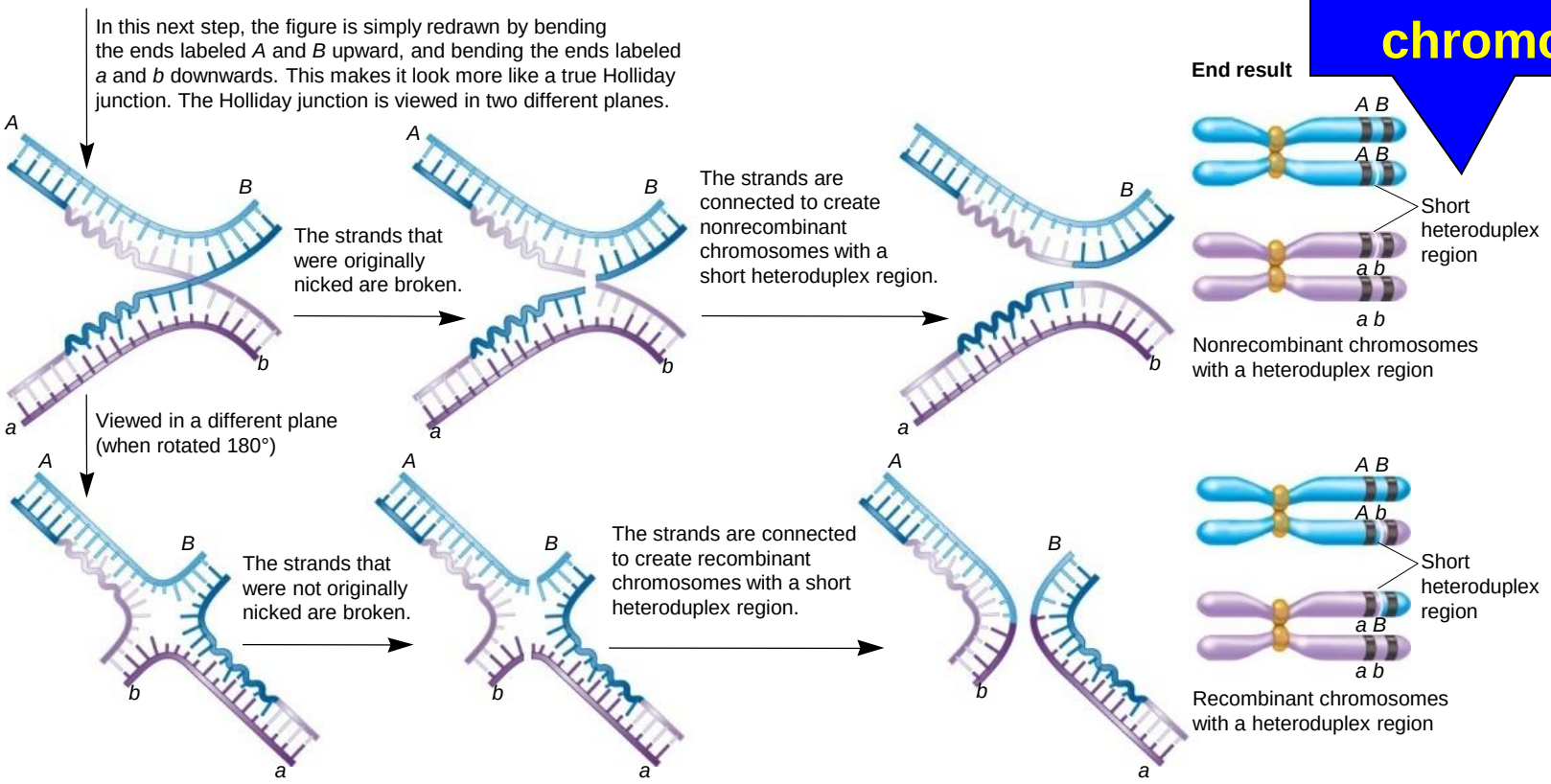
Capable of migrating in a lateral direction

Branch migration

Figure 19.2

https://www.youtube.com/watch?v=mCaFgwWH61o

Only difference between the two chromosomes



(a) The Holliday model for homologous recombination

Isomerization

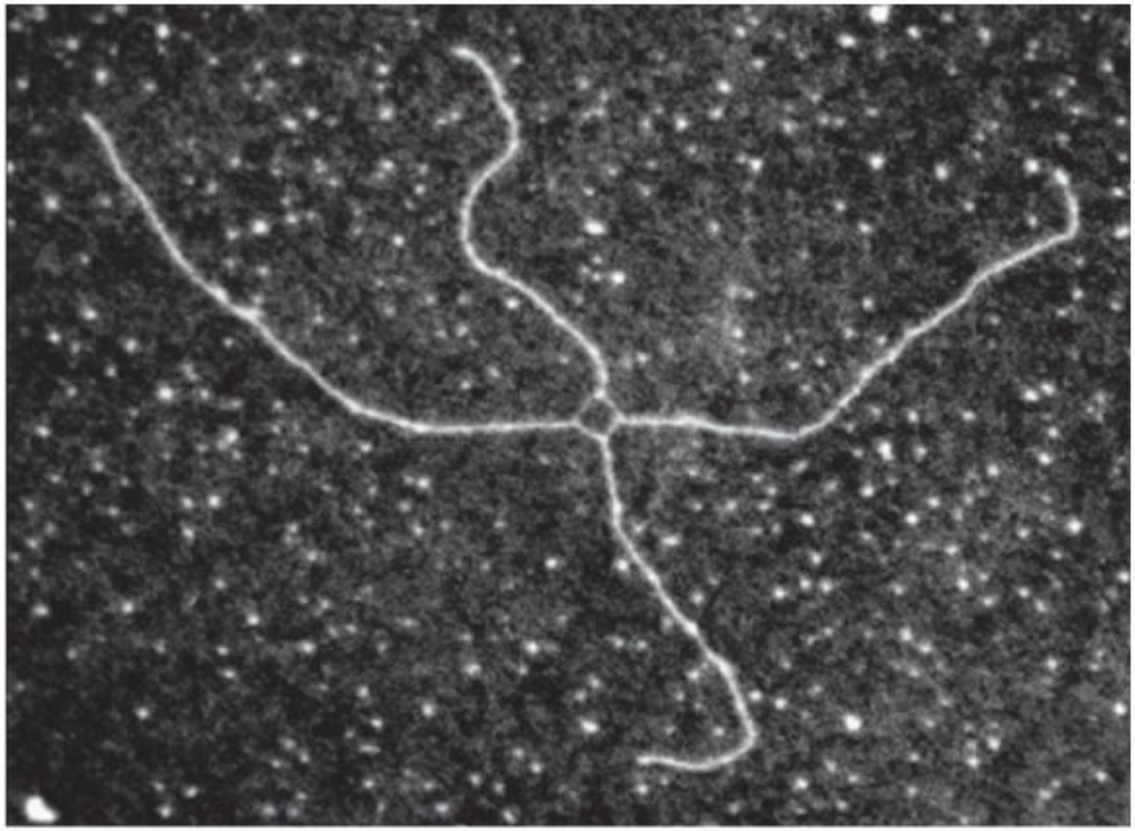
Resolution

Figure 19.2

Holliday Model of Homologous Recombination

- The Holliday model accounts for general properties of recombinant chromosomes in meiosis

- Molecular research has supported the central tenets of the Holliday model
  - A particularly convincing piece of evidence came from electron micrographs of recombination structures
    - The structure has been called a chi (χ) form
      - Its shape is similar to the Greek letter χ
      - See Figure 19.2b



(b) Micrograph of a Holliday junction

b: *From*: H. Potter & D. Dressler, "DNA recombination: in vivo and in vitro studies," *Cold Spring Harb Symp Quant Biol*. 1979, 43: 969-985. © Cold Spring Harbor Laboratory Press. Image provided by Huntington Potter, Ph.D.

Recent Models for Homologous Recombination

- Refinement of the Holliday model
  - Needed to modify the initiation phase of recombination
    - Unlikely to get nicks at the same site in each chromatid
    - Instead, it is more likely for one DNA helix (not both) to incur a single nick or a double-strand break
    - Called the **double-strand break model**
    - Refer to Figure 19.3

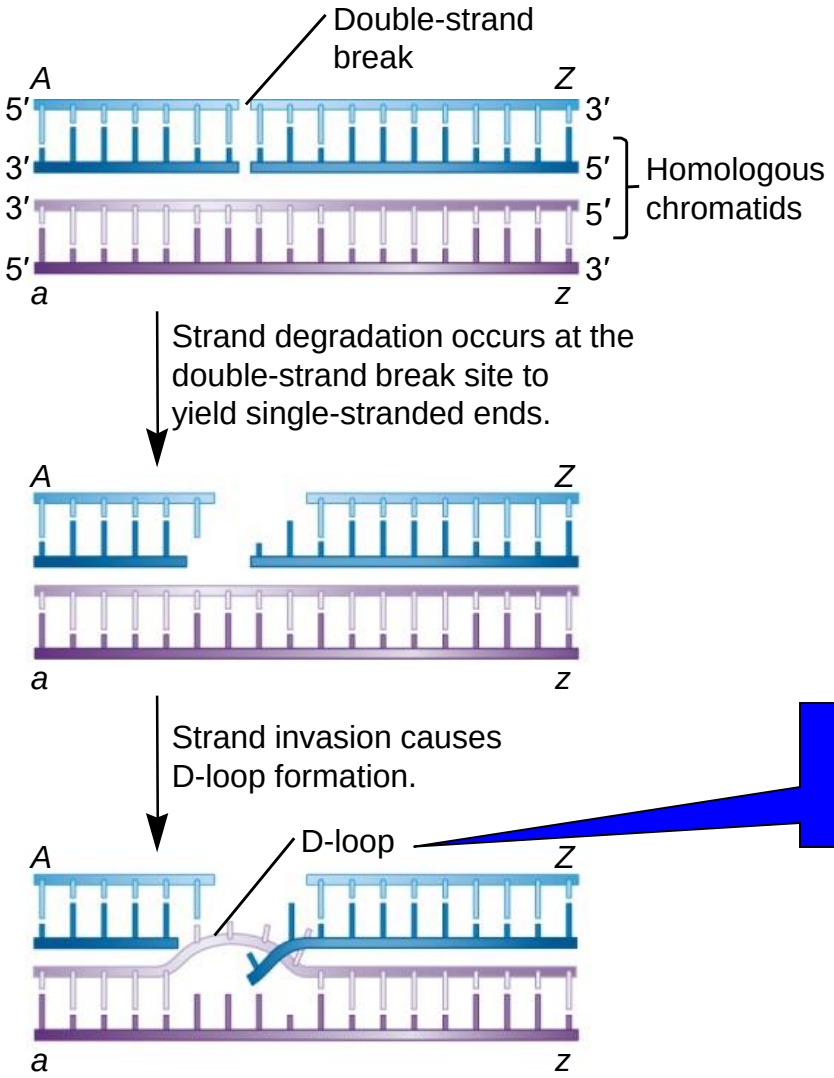
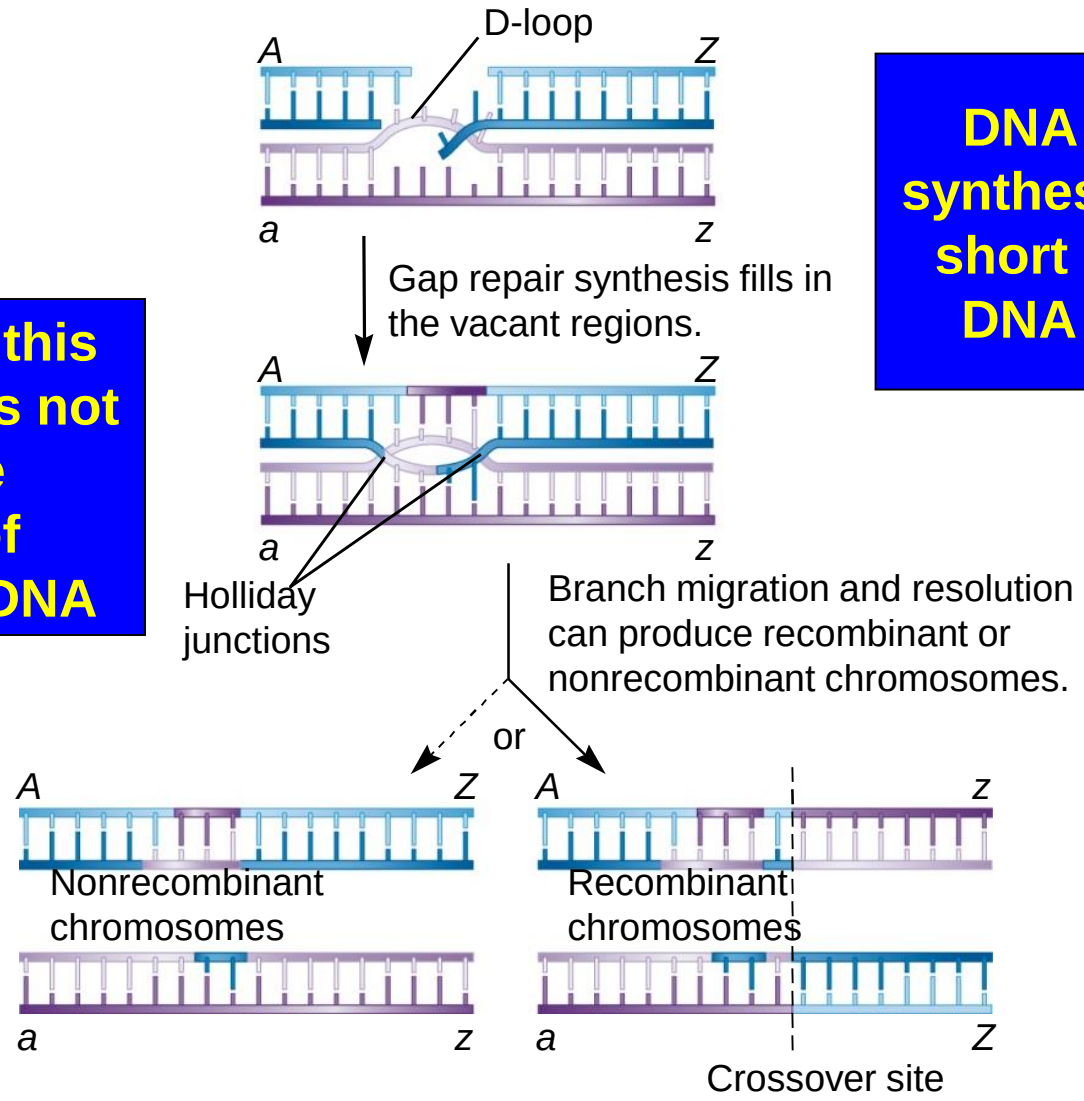


Figure 19.3

For simplicity, this illustration does not include the formation of heteroduplex DNA



DNA gap repair synthesis fills in the short gaps where DNA is missing

Figure 19.3

TABLE 19.1

E. coli Proteins That Play a Role in Homologous Recombination

Protein	Description
RecBCD	A complex of three proteins that tracks along the DNA and recognizes double-strand breaks. The complex partially degrades the double-stranded regions to generate single-stranded regions that can participate in strand invasion. RecBCD is also involved in loading RecA onto single-stranded DNA.
Single-stranded binding protein	Coats broken ends of chromosomes and prevents excessive strand degradation.
RecA	Binds to single-stranded DNA and promotes strand invasion, which enables homologous strands to find each other. It also promotes the displacement of the complementary strand to generate a D-loop.
RuvABC	This protein complex binds to Holliday junctions. RuvAB promotes branch migration. RuvC is an endonuclease that cuts the crossed or uncrossed strands to resolve Holliday junctions into separate chromosomes.
RecG	RecG protein can also promote branch migration of Holliday junctions.

Table 19.1

Gene Conversion

- Homologous recombination can lead to **gene conversion**
  - Two different alleles become identical

- Gene conversion occurs in one of two ways
  1. **DNA mismatch repair**
     - Refer to Figure 19.4
  2. **DNA gap repair synthesis**
     - Refer to Figure 19.5

Gene conversion by DNA mismatch repair

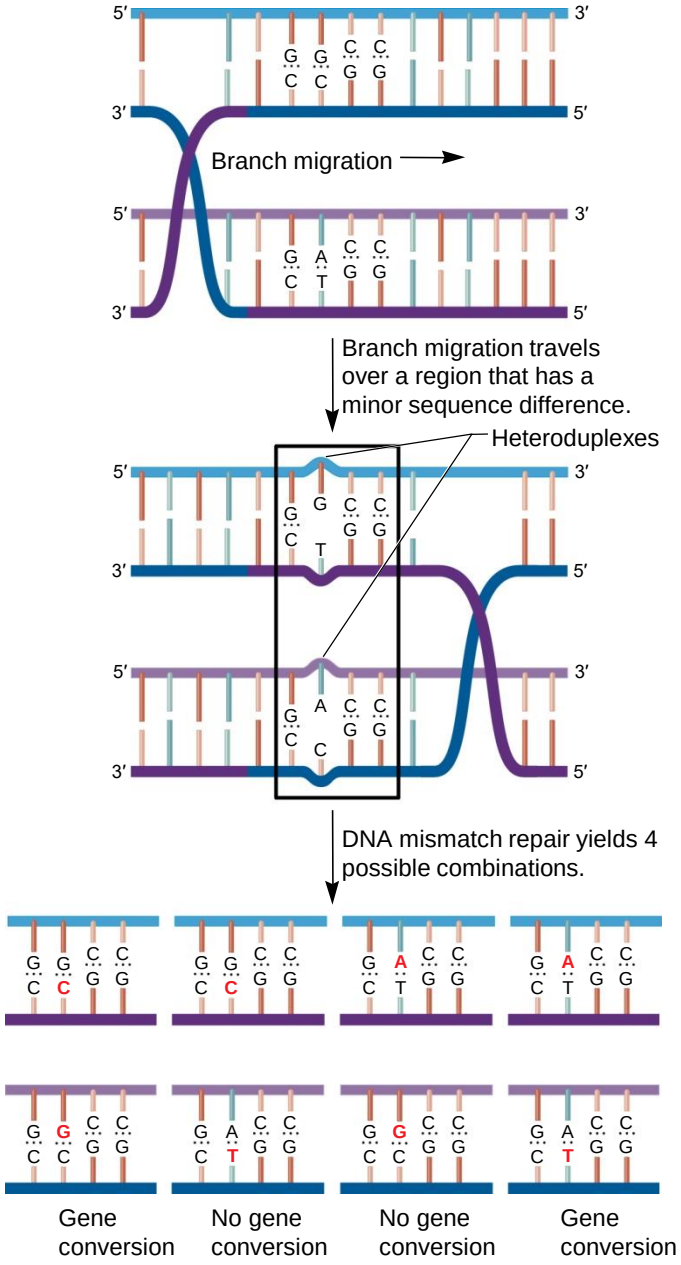


Figure 19.4

Gene conversion by gap repair synthesis

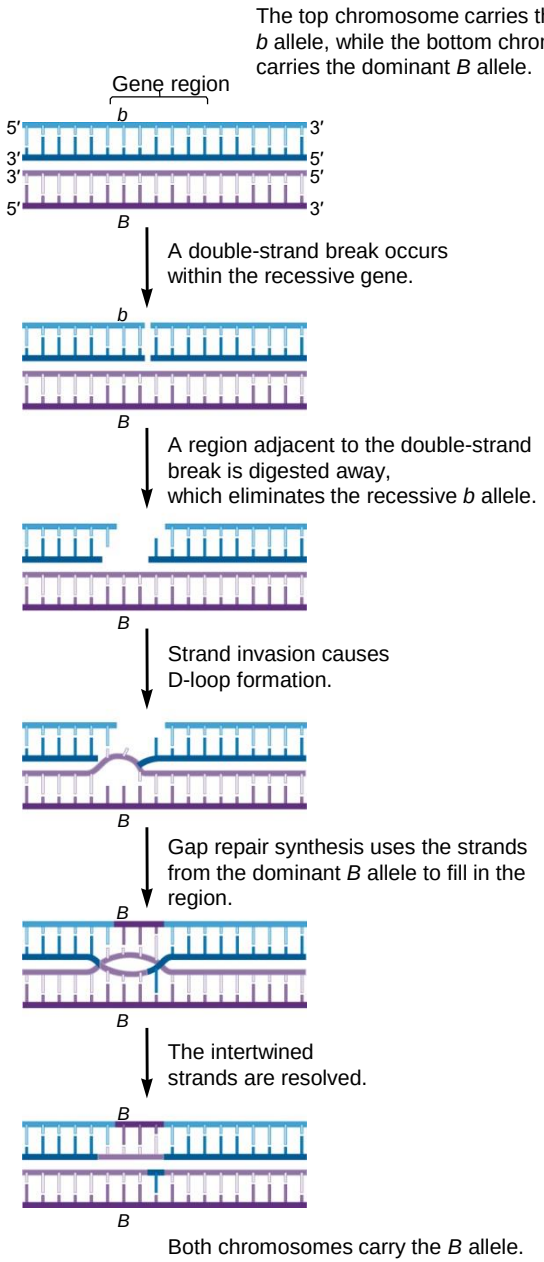


Figure 19.5

19.2 Immunogenetics

- The structure and function of antibodies
- How antibody diversity is generated in vertebrates via site-specific recombination

Immunogenetics

- **Antibodies** or **immunoglobulins (Igs)** are produced by specific cells of the immune system known as lymphocytes

- B cells produce antibodies in mammals
  - Antibodies bind to specific **antigens**
    - Proteins or other molecules that are foreign to the individual

- Each B cell only produces one specific antibody, and a single individual can produce millions of B cells

- Many different lymphocytes are necessary to produce all of the antibodies necessary for the immune system
  - System of antibody generation must produce many different proteins

- Diversity is created by
  - **Site-specific recombination**
  - **Imprecise end-joining**
  - **Somatic hypermutation**

Antibody Diversity

- If a separate gene was needed for each distinct antibody, the genome would need millions of antibody genes!

- Instead, antibody diversity is created with an unusual mechanism – **site-specific recombination**
  - Antibody genes are cut and reconnected in many different ways
  - Only a few genes can produce millions of antibodies

- **Antibody structure**
  - Antibodies are tetrameric proteins
    - Two identical **heavy chains**
      - **Constant region** may be common between different antibody molecules
      - **Variable region** with shared and unique sequences
    - Two identical **light chains**
      - **Constant regions** with shared sequences
      - **Variable regions** with shared and unique sequences
    - Refer to Figure 19.6

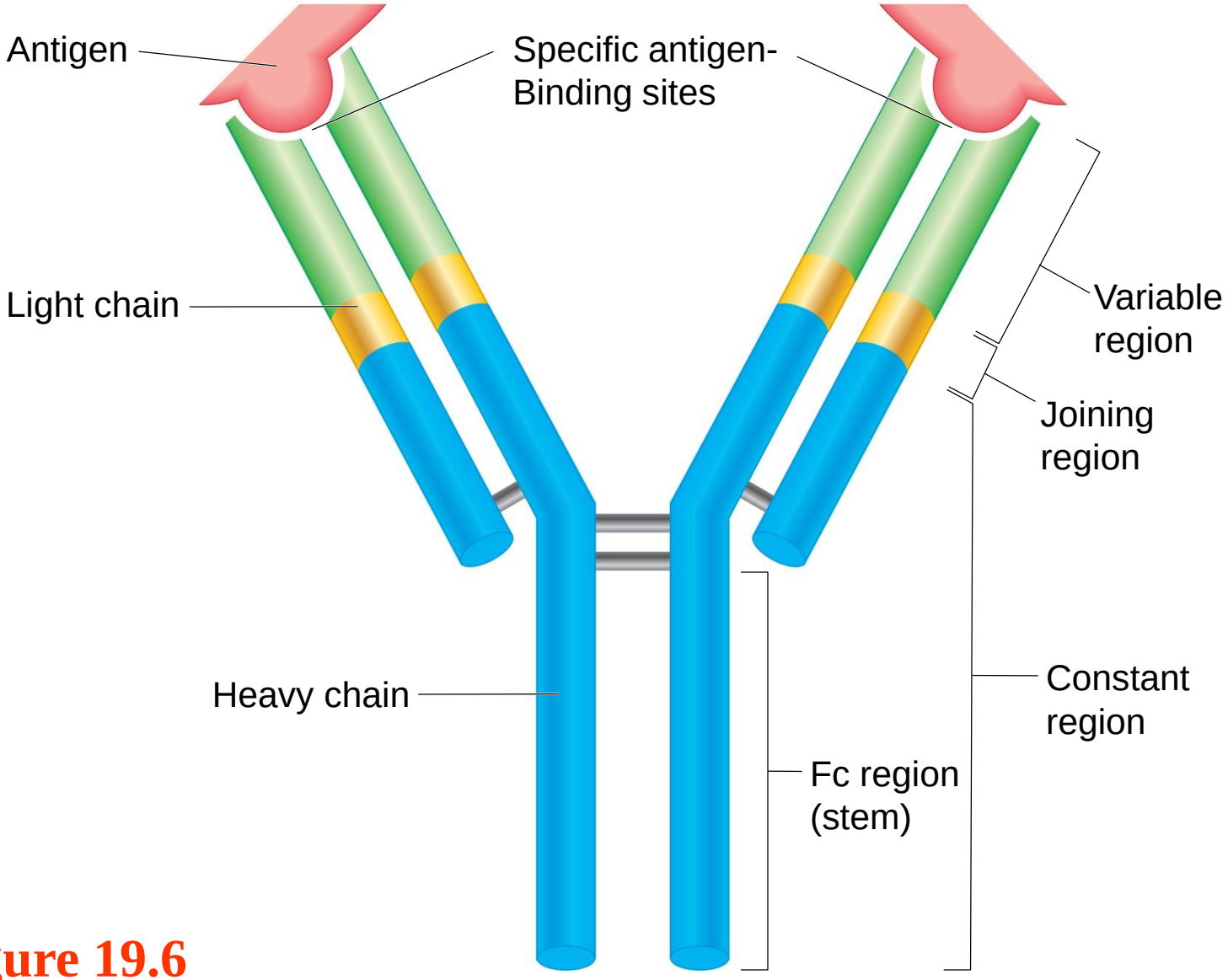


Figure 19.6

Site-Specific Recombination

- Mechanism of site-specific recombination

1. **Recombination signal sequence**
   - At the end of every V domain and beginning of every J domain
   - **RAG1** and **RAG2** proteins recognize the recombination signal sequence and make two double-strand breaks in the DNA

2. Intervening DNA is lost, and the ends are joined by **nonhomologous end-joining (NHEJ) proteins**

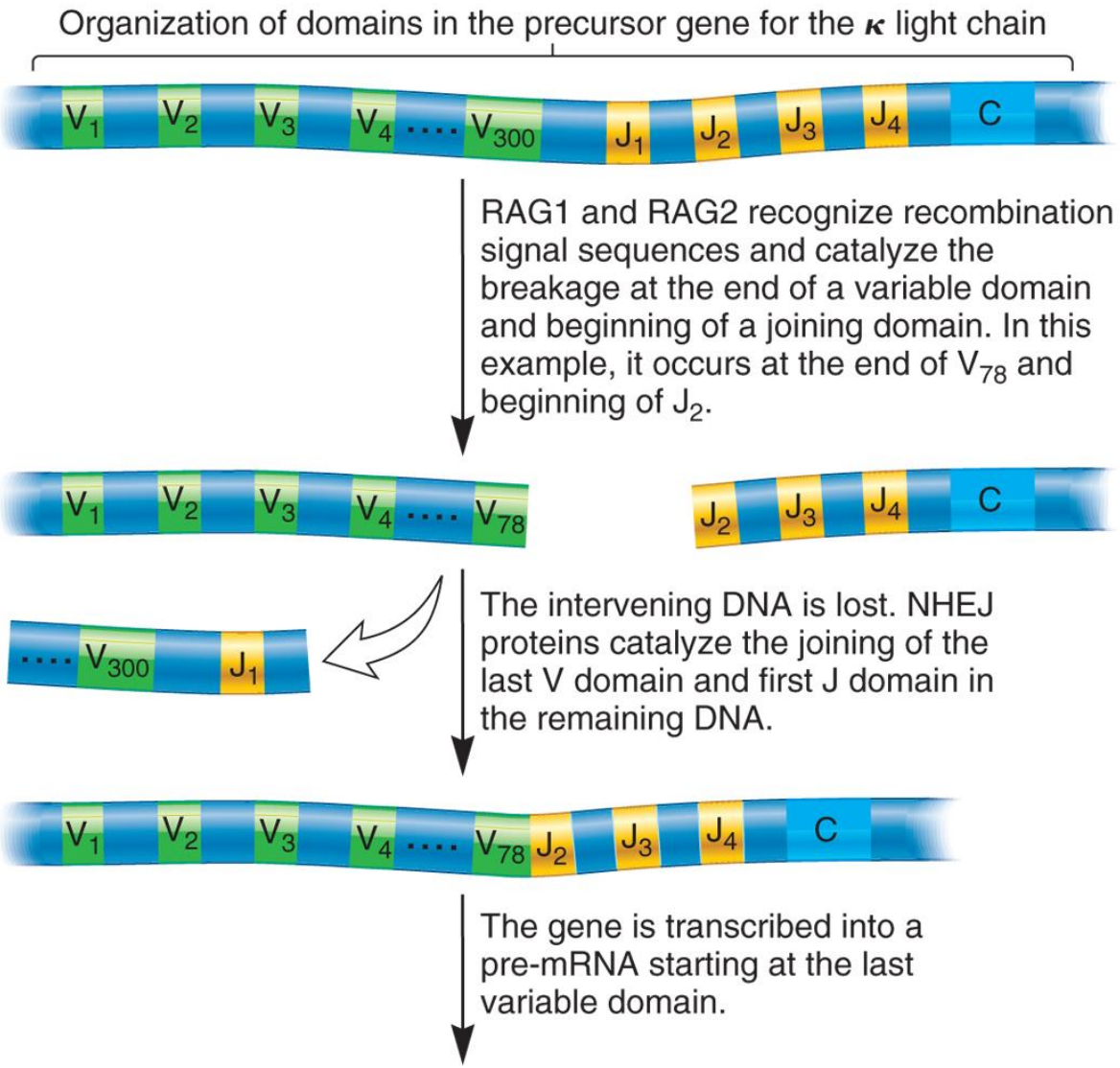

Organization of domains in the precursor gene for the κ light chain
V1 V2 V3 V4 .... V300 J1 J2 J3 J4 C
RAG1 and RAG2 recognize recombination signal sequences and catalyze the breakage at the end of a variable domain and beginning of a joining domain. In this example, it occurs at the end of V78 and beginning of J2.
V1 V2 V3 V4 .... V78
J2 J3 J4 C
.... V300 J1
The intervening DNA is lost. NHEJ proteins catalyze the joining of the last V domain and first J domain in the remaining DNA.
V1 V2 V3 V4 .... V78 J2 J3 J4 C
The gene is transcribed into a pre-mRNA starting at the last variable domain.

Figure 19.7

3. After transcription, the J and C domains are connected by splicing of the pre-mRNA

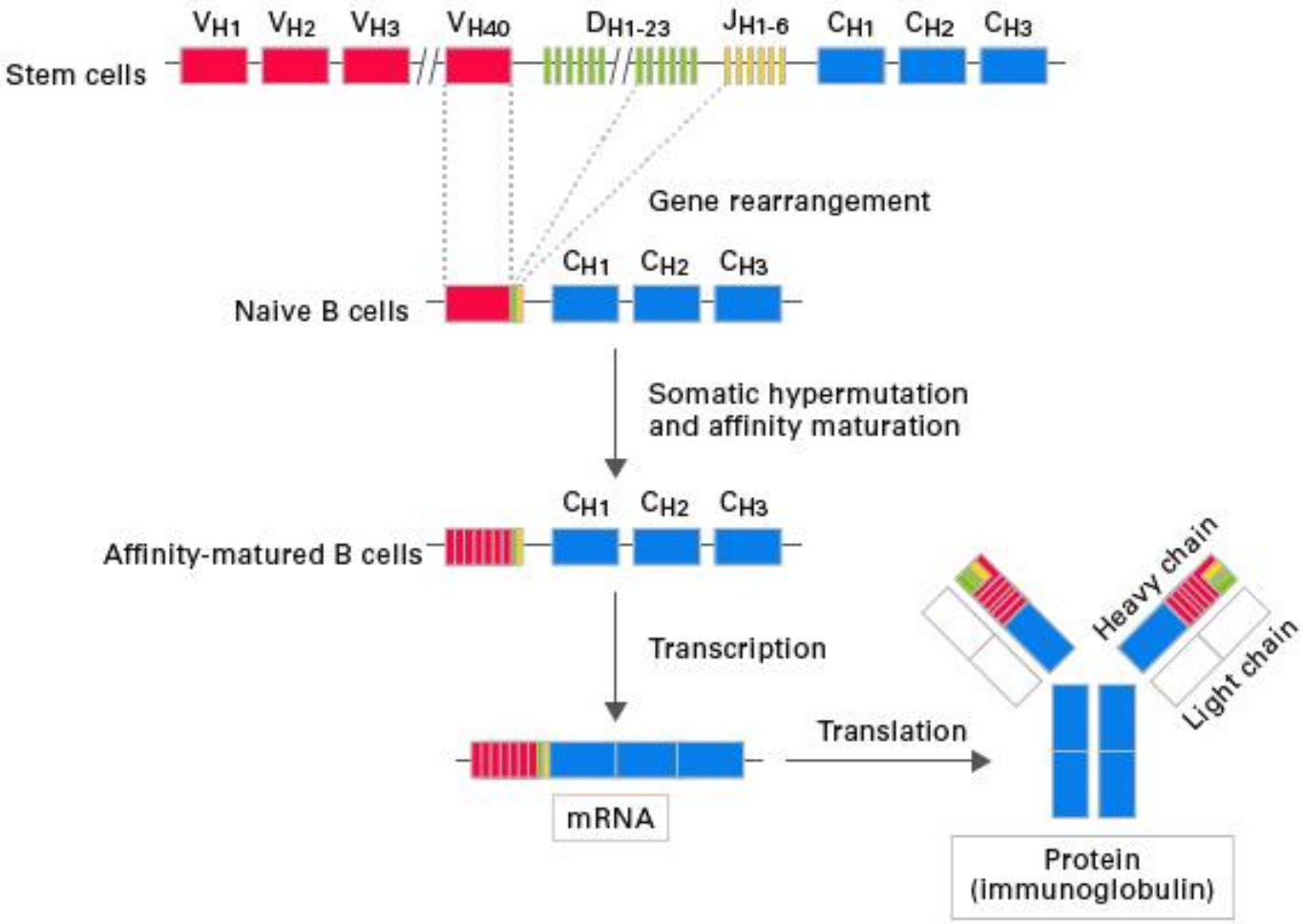
4. The mature mRNA is translated, producing a light chain with one V, one J, and one C domain (each randomly chosen in each B cell)

5. The light chains assemble with the heavy chains to produce the functional antibody

– Refer to Fig. 19.7

- Process is called **V(D)J recombination**
  - D is found only in the heavy-chain genes

Heavy chain


V_{H1} V_{H2} V_{H3} V_{H40} D_{H1-23} J_{H1-6} C_{H1} C_{H2} C_{H3}
Stem cells
Gene rearrangement
C_{H1} C_{H2} C_{H3}
Naive B cells
Somatic hypermutation
and affinity maturation
C_{H1} C_{H2} C_{H3}
Affinity-matured B cells
Transcription
Translation
mRNA
Heavy chain
Light chain
Protein
(immunoglobulin)

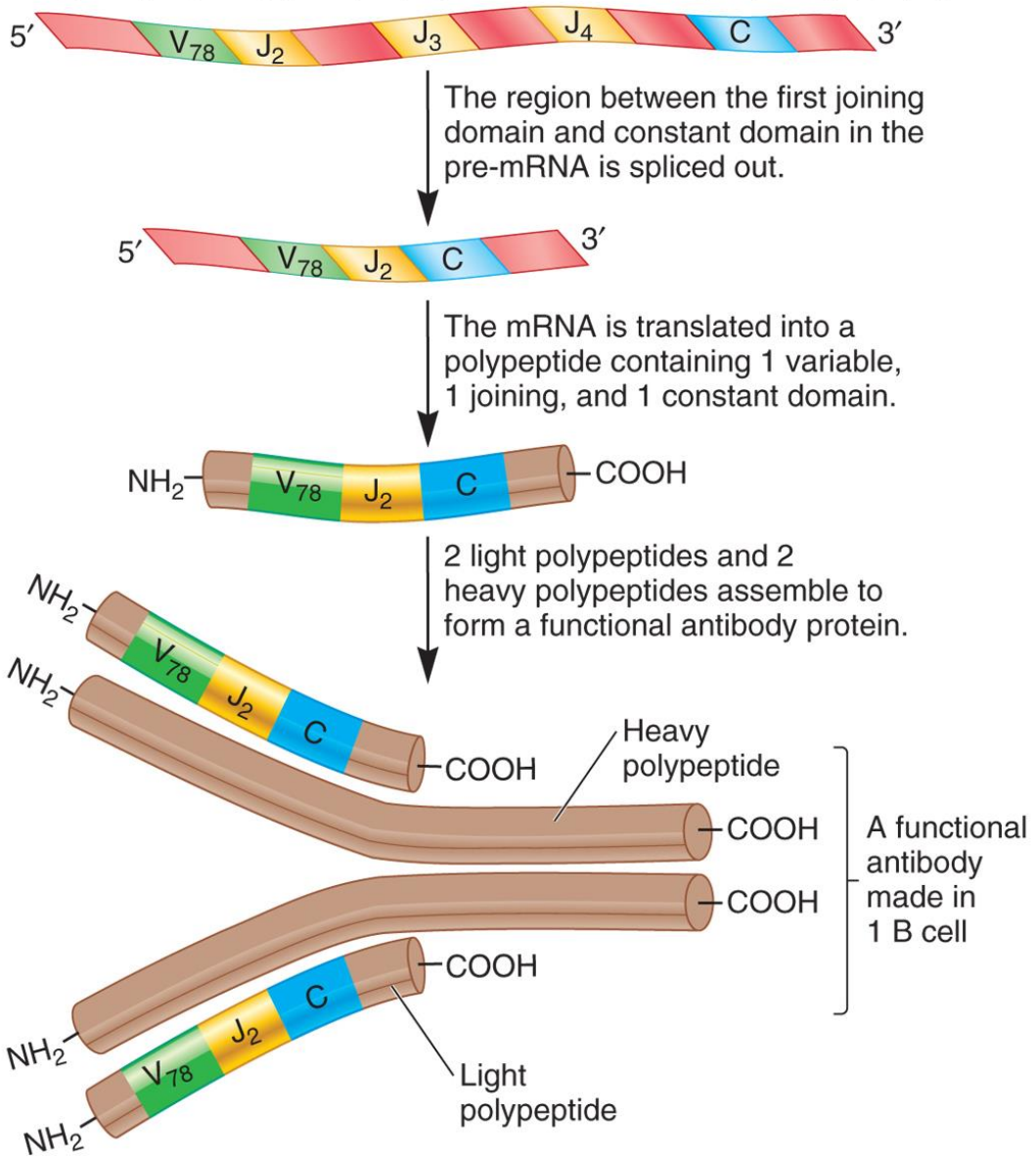


Figure 19.7

- Possible antibody diversity generated by V(D)J recombination is staggering

  - 300 different V sequences
  - 4 different J sequences
    - » 1200 possible light chains

  - 500 different V sequences
  - 12 different D sequences
  - 4 different J sequences
    - » 24,000 possible heavy chains

    - » 28,800,000 possible antibody molecules!

Imprecise End-Joining

- Second mechanism for antibody diversity

- In antibody gene rearrangement process, connection is catalyzed by NHEJ proteins
  - These are not entirely precise
    - They may add or lose a few nucleotides between V and J domains
  - Adds further diversity

Somatic Hypermutation

- Third mechanism for antibody diversity
- Antibody genes have a high rate of mutation in B cells
  - Cytosines are often deaminated into uracils, leading to C to T mutations
  - Also, the uracils recuit lesion-replicating DNA polymerases, which are error-prone
- This domatic hypermutations add even more diversity

https://www.ibiology.org/immunology/b-cell-development/
https://www.youtube.com/watch?v=YYZvT5qneEs&feature=emb_rel_pause

19.3 Transposition

- The organization of sequences within different types of transposable elements
- How transposons and retroelements move to new locations in a genome
- The effects of transposable elements on gene function

Transposition

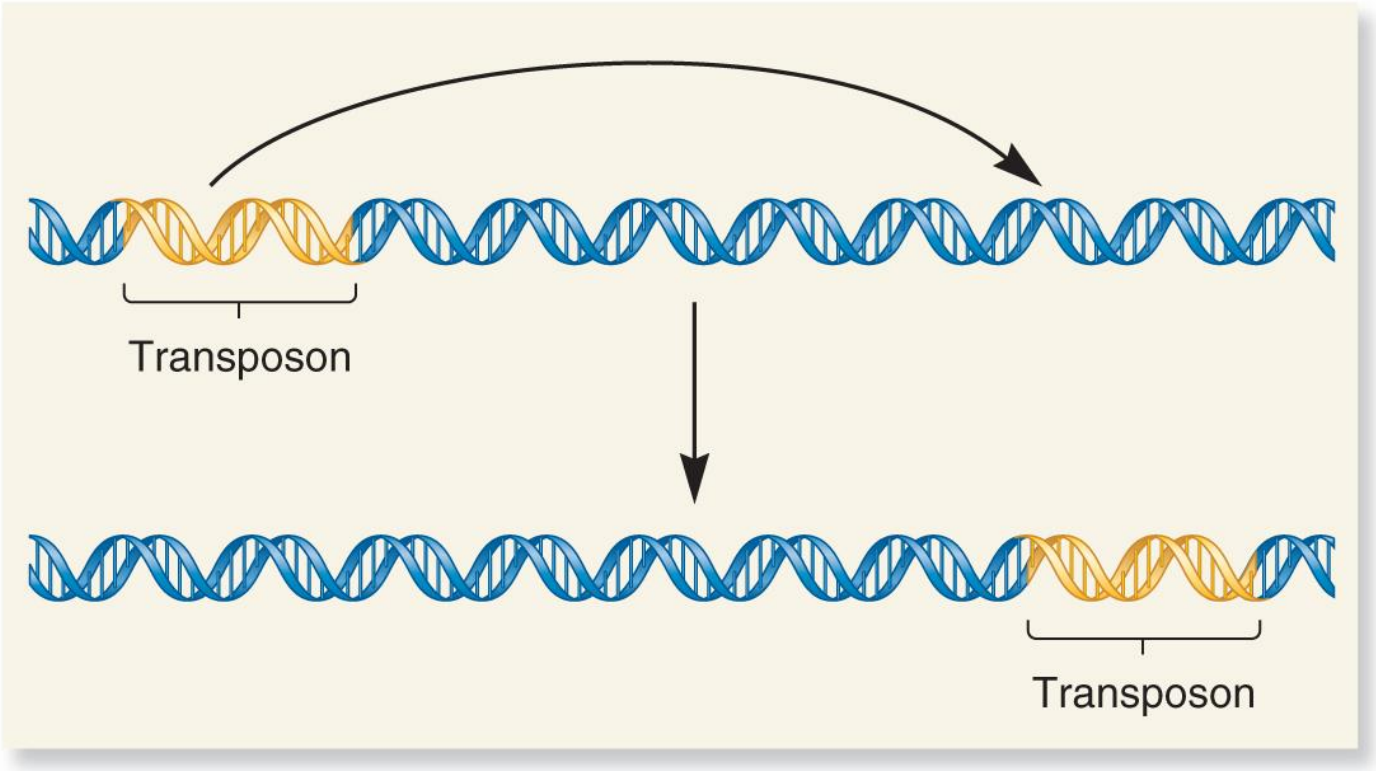
- Transposition involves the integration of small segments of DNA into the chromosome
  - Can occur at many different locations within the genome

- The DNA segments are **transposable elements (TEs)**
  - Sometimes referred to as “jumping genes”

- TEs were first identified by Barbara McClintock in the early 1950s from her classical studies in corn
  - Since then, many different types of TEs have been found in species as diverse as bacteria, fungi, plants and animals

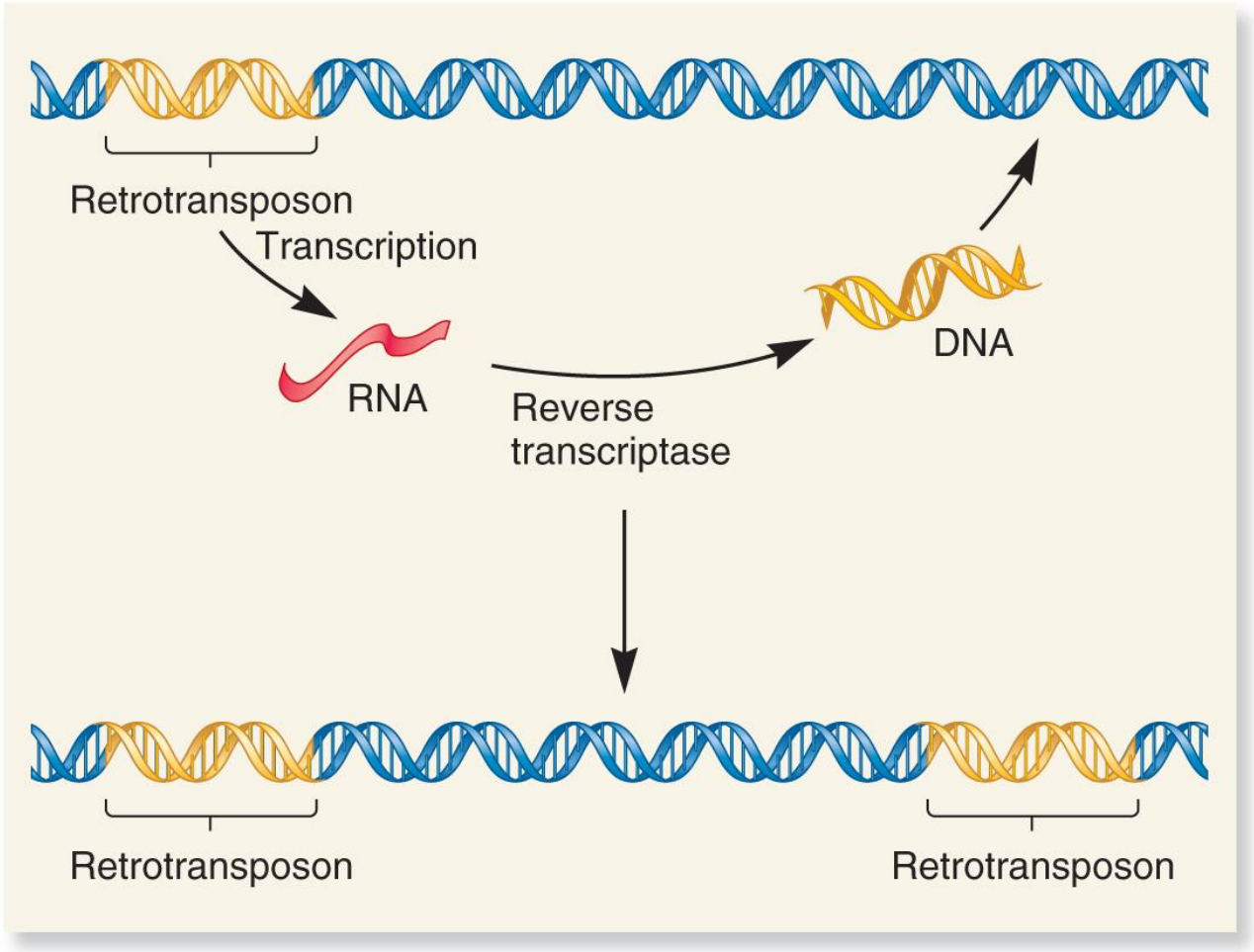
Transposition Pathways

- Two basic transposition pathways have been identified
  - **Simple transposition**
    - The TE is removed from original site and transferred to a new site by a “cut and paste” mechanism
    - These TEs are called **transposons**
  - **Retrotransposition**
    - The TE moves via **retrotransposition** – TE is transcribed into RNA, then reverse transcriptase makes a second copy in DNA
    - These TEs are called **retrotransposons** or **retroelements**
  - Refer to Figure 19.8



(a) Simple transposition

Figure 19.8a



(b) Retrotransposition

Figure 19.8b

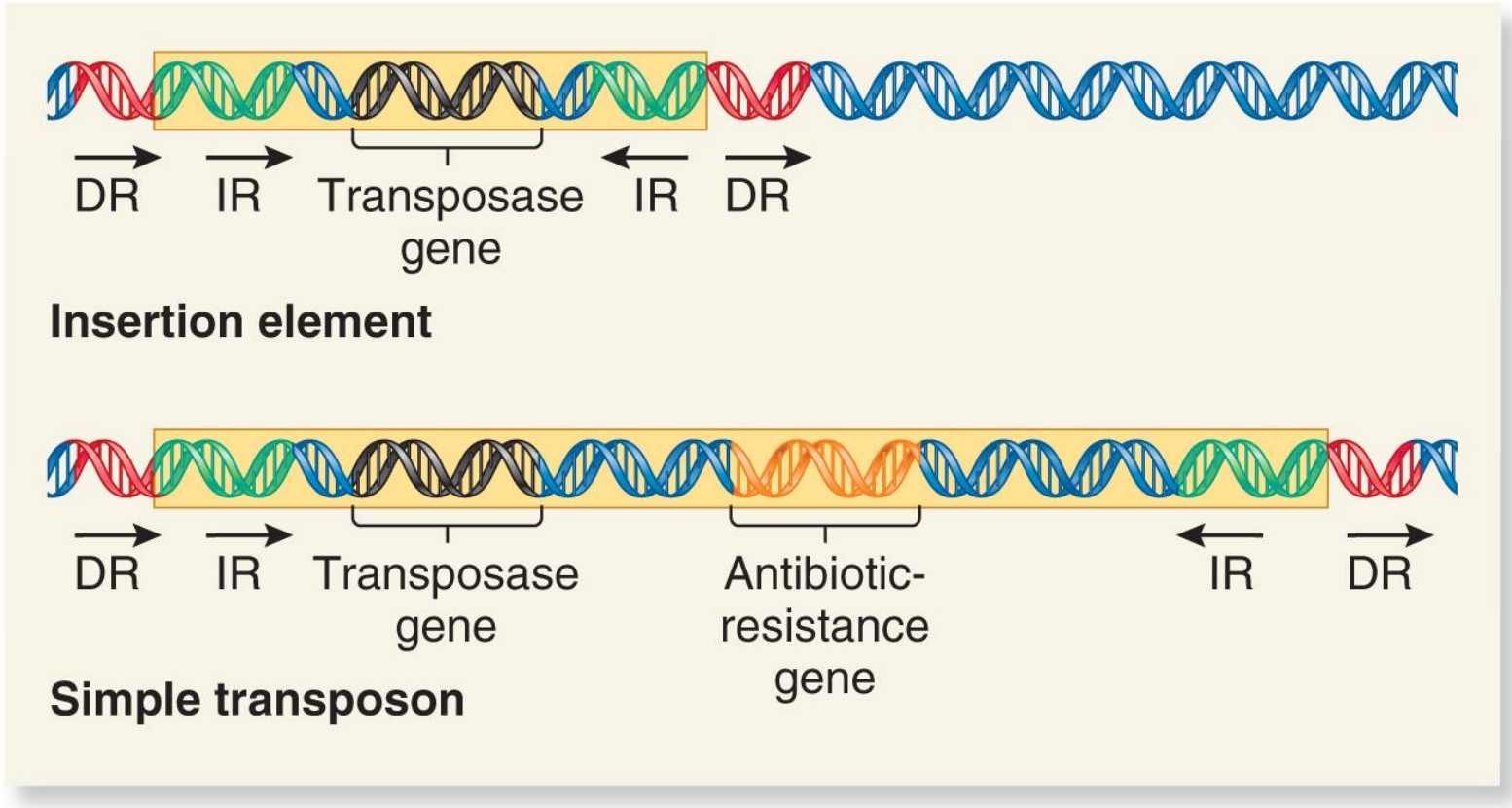
Different Types of Transposable Elements Have Characteristic DNA Sequences

- All TEs are flanked by **direct repeats (DRs)**, also called **target-site duplications**
  - These direct repeat sequences go in the same direction on both sides

- **Insertion Elements (IS element)**
  - The simplest type of TE
  - Both ends have **inverted repeats (IRs)**
    - Ranging from 9-40 bp long
  - Also contains the ***transposase*** gene

- **Simple Transposons**
  - In addition to
    - Flanking direct repeats
    - Inverted repeats
    - Transposase gene
  - The simple transposon also carries one or more genes not necessary for transposition
    - Example: Antibiotic resistance gene

- Refer to Figure 19.9a



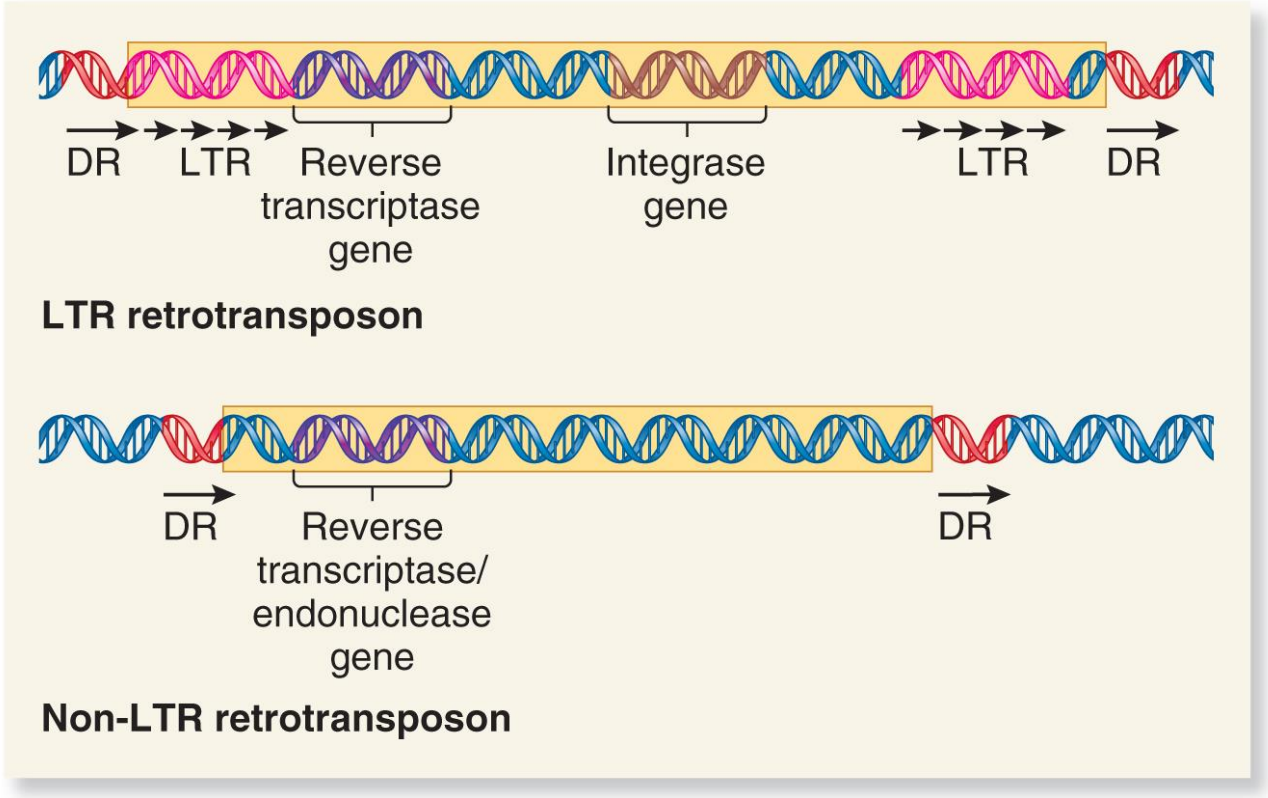
(a) Elements that move by simple transposition

Figure 19.9a

- **LTR Retrotransposons**
  - Related to retroviruses
    - Move around the genome, but cannot produce viral particles
  - Contain long terminal repeats (LTRs) at both ends
    - A few hundred bp long
  - Encode reverse transcriptase and integrase

- **Non-LTR Retrotransposons**
  - Less like retroviruses
  - May encode reverse transcriptase /endonuclease
  - Some derived from host genes
    - Ex: ***Alu*** in humans

- Refer to Figure 19.9b



(b) Elements that move by retrotransposition (via an RNA intermediate)

Figure 19.9b

- TEs are complete (or **autonomous**) when they contain all the information necessary for transposition to occur

- TEs are incomplete (or **nonautonomous**) when they lack a gene that is necessary for transposition to occur

- Example: In corn
    - *Ac* or Activator element is autonomous
    - *Ds* element is nonautonomous - it lacks transposase and needs transposase from *Ac* element to move

Transposase

- The enzyme transposase catalyzes the removal of a TE and the its reinsertion at another location
- Transposase recognize the inverted repeats at the ends of a TE and bring them closer together
- Refer to Figure 19.10

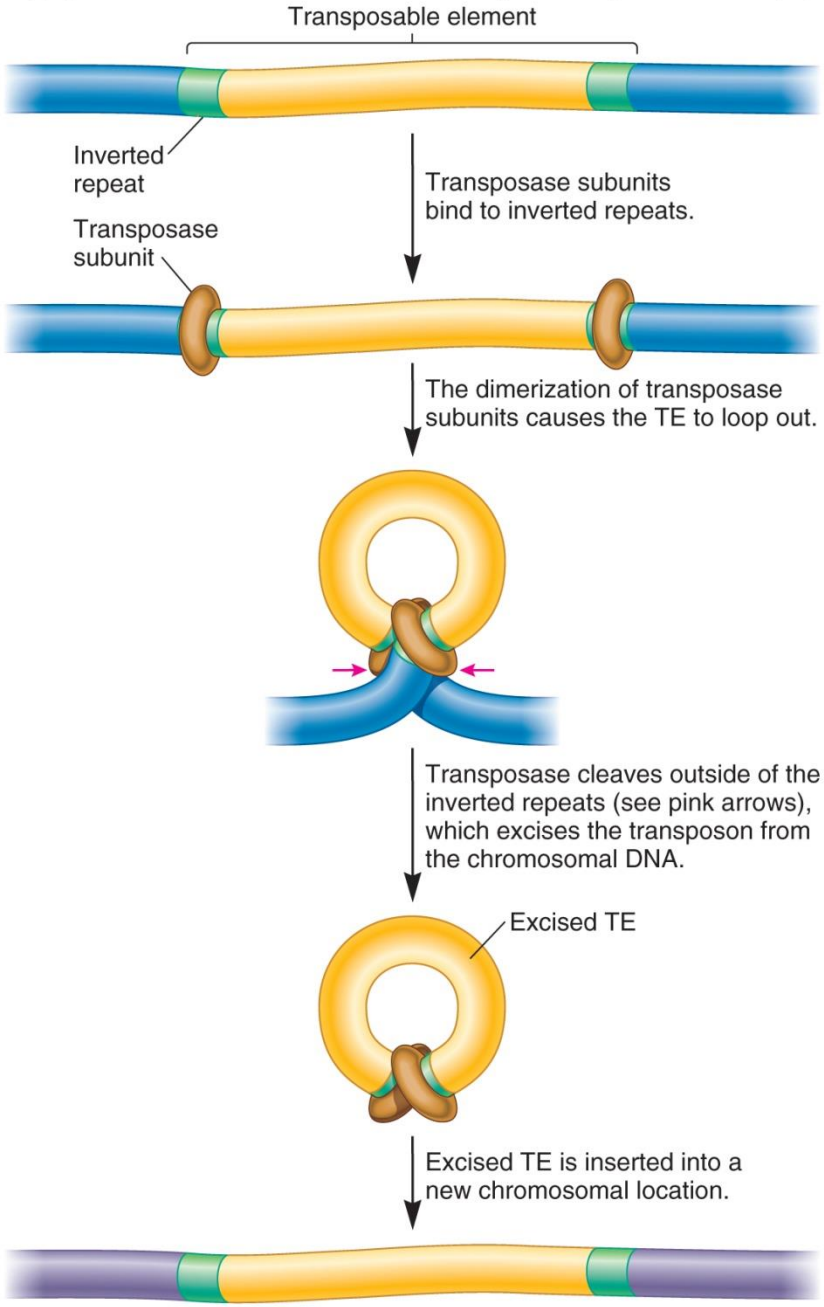
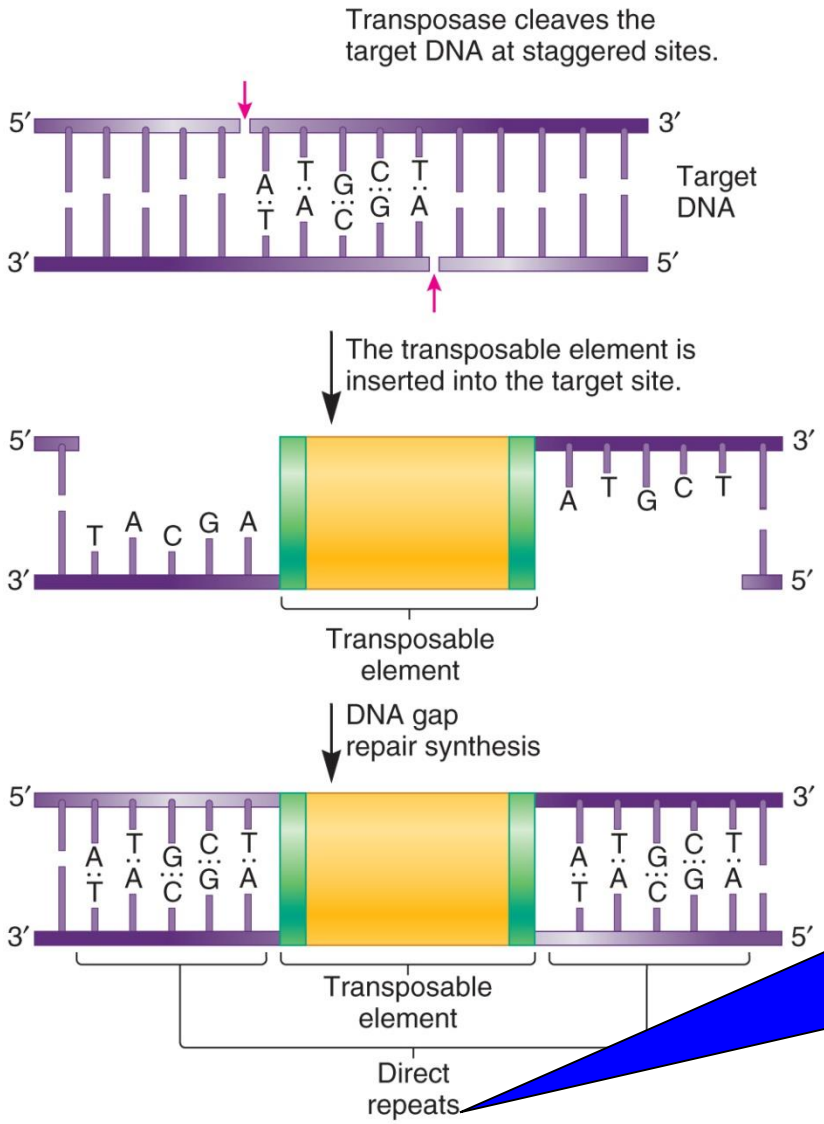


Figure 19.10a

(a) Movement of a transposon via transposase



(b) The formation of direct repeats

They are in the same direction and are repeated at both ends of the element

Figure 19.10b

- Even simple transposition can increase the number of transposons in the genome

- Transposition often occurs around time of replication
  - One of these TEs can transpose ahead of the fork where it is copied again.
  - One genome will still have one TE, but the other will now have two copies

- Refer to Figure 19.11

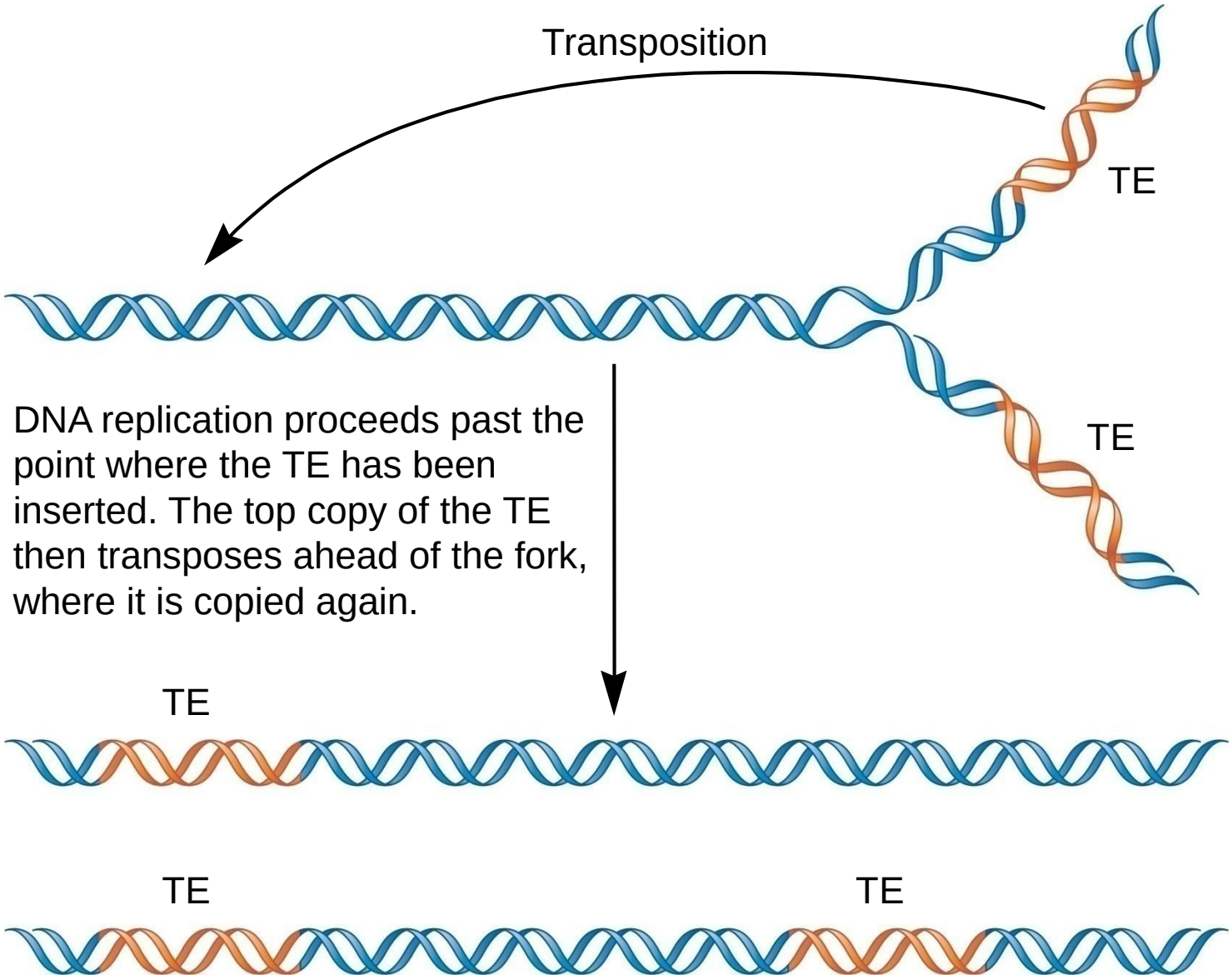


Figure 19.11

Reverse Transcriptase

- Retroelements use an RNA intermediate in their transposition mechanism

- The movement of retroelements also requires two key enzymes:
  - **Reverse transcriptase**
  - **Integrase**

- Refer to Figure 19.12

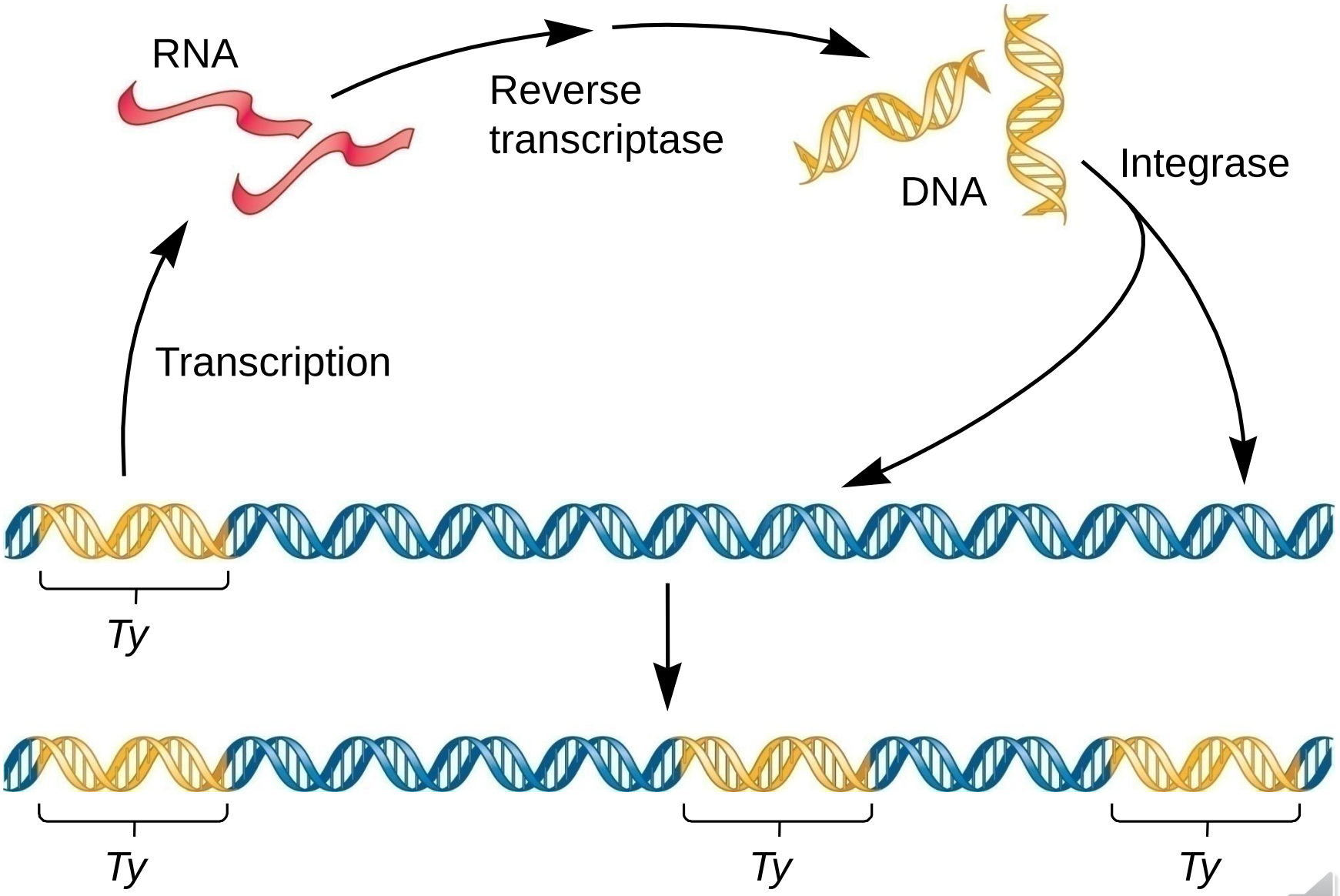


Figure 19.12

Non-LTR retrotransposons

- Current model is **target-site primed reverse transcription (TPRT)**

1. Retrotransposon transcribed into RNA with a poly-A tail, and the target DNA is nicked
2. The RNA binds to the nicked site by AT base pairing
3. Reverse transcriptase makes a DNA copy of the RNA
4. Endonuclease makes a second cut nearby
5. Retrotransposon DNA is ligated into the target site

– Refer to Figure 19.13

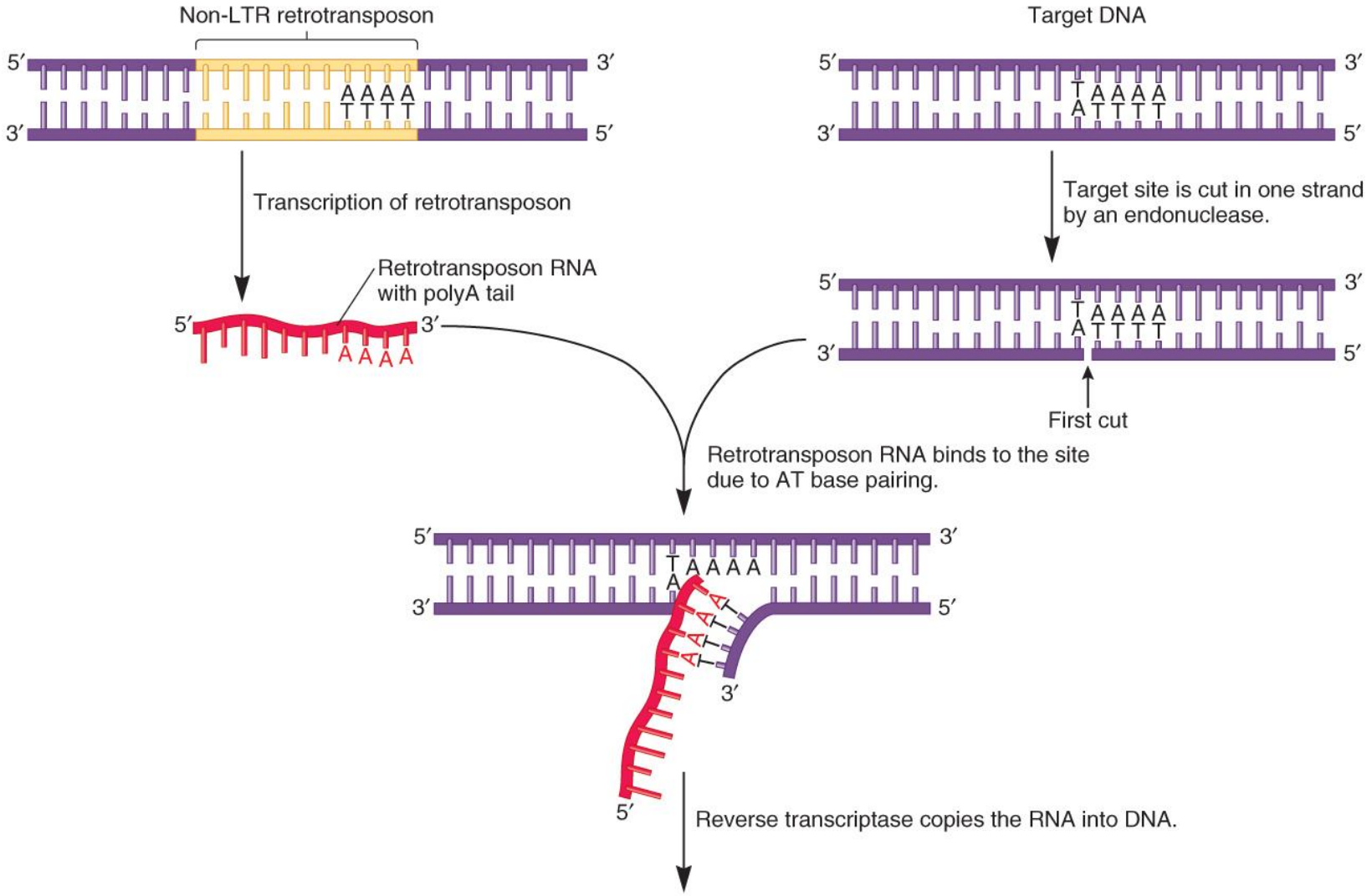


Figure 19.13

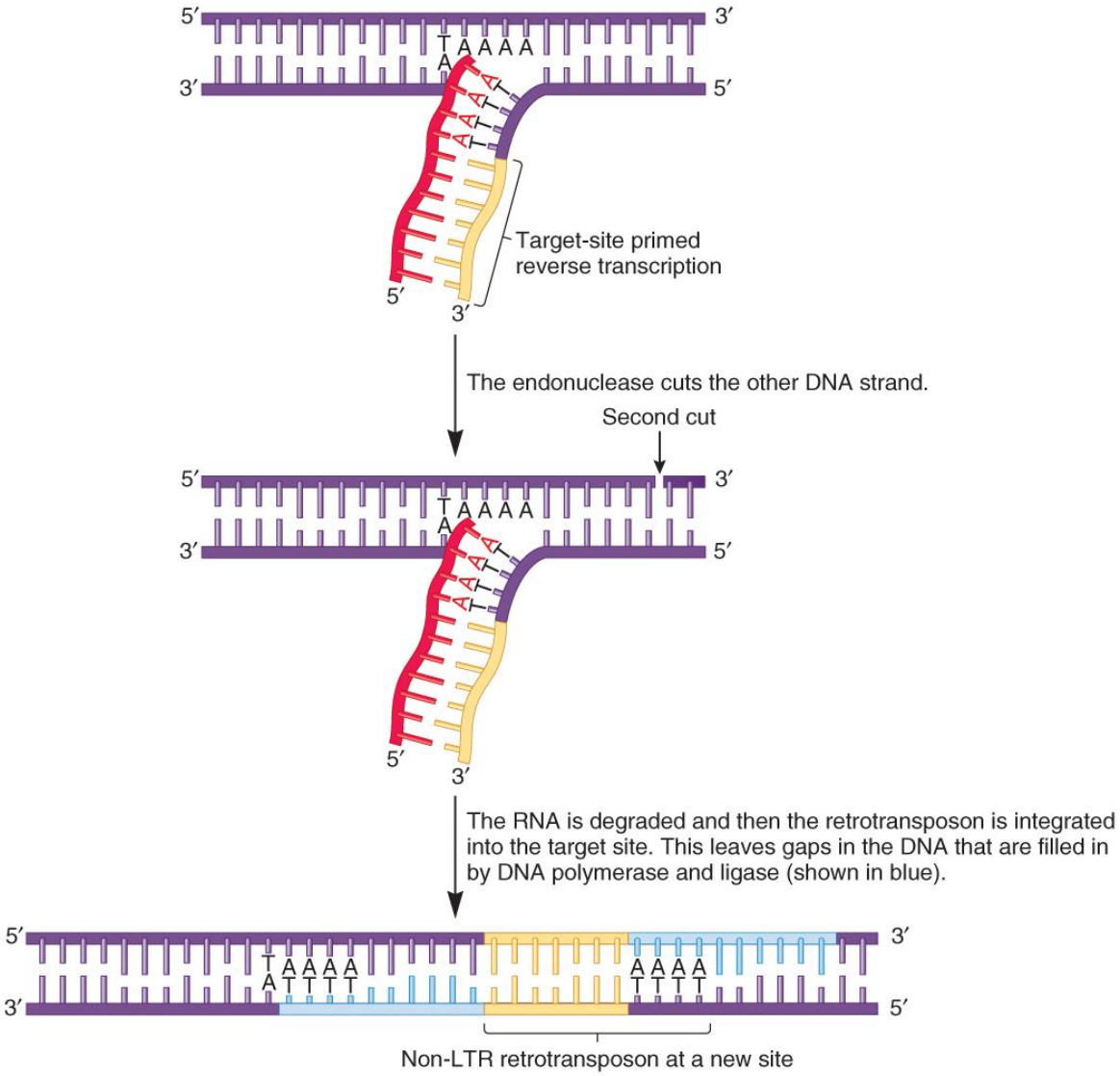


Figure 19.13

Transposable Elements Influence on Mutation and Evolution

- Over the past few decades, researchers have found that transposable elements probably occur in the genomes of all species
- Can rapidly enter genome and proliferate
- Many different kinds
- Can be a major part of genome

- In mammals, for example
  - **LINEs**
    - **Long interspersed elements**
    - Usually 1,000 to 10,000 bp long
    - Found in a few thousand to several hundred thousand copies per genome
  - **SINEs**
    - **Short interspersed elements**
    - Less than 500 bp in length
    - Example: *Alu* sequence
      - About 1 million copies in the human genome
      - 10% of the genome!

TABLE 19.2

Examples of Transposable Elements

Element	Type	Approximate Length (bp)	Description
Bacterial			
IS1	Transposon	768	An insertion element that is commonly found in five to eight copies in *E. coli*.
Tn10	Transposon	9300	One of many different bacterial transposons that carries antibiotic resistance.
Tn951	Transposon	16,600	A transposon that provides bacteria with genes that allow them to metabolize lactose.
Yeast			
Ty elements	Retrotransposon	6300	Found in *S. cerevisiae* at about 35 copies per genome.
Drosophila			
P elements	Transposon	500–3000	A transposon that may be found in 30–50 copies in P strains of *Drosophila*. It is absent from M strains.
Copia-like elements	Retrotransposon	5000–8000	A family of *copia*-like elements found in *Drosophila*, which vary slightly in their lengths and sequences. Typically, each family member is found at about 5–100 copies per genome.
Humans			
Alu sequence	Retrotransposon	300	A SINE that is abundantly interspersed throughout the human genome.
L1	Retrotransposon	6500	A LINE found in about 500,000 copies in the human genome.
Plants			
Ac / *Ds*	Transposon	4500	*Ac* is an autonomous transposon found in corn and other plant species. It carries a transposase gene. *Ds* is a nonautonomous version that lacks a functional transposase gene.
Opie	Retrotransposon	9000	A retrotransposon found in plants that is related to the *copia*-like elements found in animals.

Table 19.2

TABLE 19.3

Abundance of TEs in the Genomes of Selected Species

Species	Percentage of the Total Genome Composed of Transposable Elements*
Frog (*Xenopus laevis*)	77
Corn (*Zea mays*)	60
Human (*Homo sapiens*)	45
Mouse (*Mus musculus*)	40
Fruit fly (*Drosophila melanogaster*)	20
Nematode (*Caenorhabditis elegans*)	12
Yeast (*Saccharomyces cerevisiae*)	4
Bacterium (*Escherichia coli*)	0.3

*In some cases, the abundance of TEs may vary somewhat among different strains of the same species. The values reported here are typical values.

Table 19.3

- The biological significance of transposons in evolution remains a matter of debate

  - **Selfish DNA hypothesis**
    - TEs exist because they can
    - Like parasites, can proliferate in host as long as they do not overly harm the host

  - **TEs offer an advantage**
    - In bacterial, may carry antibiotic-resistance genes
    - May cause insertion of exons into the coding region of other genes, providing new functions – **exon shuffling**

- Transposons have a variety of effects on chromosome structure and gene expression
  - Insertions can cause **mutations**
    - These play a role in diseases such as hemophilia, muscular dystrophy, breast cancer, and colon cancer
  - When transposon activity is not regulated and kept under control, they can cause chromosomal abnormalities and sterility
    - Example: *Drosophila* crosses that introduce P elements to a naive strain
    - Known as **hybrid dysgenesis**

TABLE 19.4

Possible Consequences of Transposition

Consequence	Cause
Chromosome Structure	
Chromosome breakage	Excision of a TE.
Chromosomal rearrangements	Homologous recombination between TEs located at different positions in the genome.
Gene Expression	
Mutation	Incorrect excision of TEs.
Gene inactivation	Insertion of a TE into a gene.
Alteration in gene regulation	Transposition of a gene next to regulatory sequences or the transposition of regulatory sequences next to a gene.
Alteration in the exon content of a gene	Insertion of exons into the coding sequence of a gene via TEs. This phenomenon is called exon shuffling.
Gene duplications	Insertion of a gene into a transposon that transposes to another site in the genome.

Table 19.4

https://www.youtube.com/watch?v=bdnqpoeIs6A