Fundamentals of DNA, chromosomes, and cells

CONTENTS

1.1 THE STRUCTURE AND FUNCTION OF NUCLEIC ACIDS 2

1.2 THE STRUCTURE AND FUNCTION OF CHROMOSOMES 8

1.3 DNA AND CHROMOSOMES IN CELL DIVISION AND THE CELL CYCLE 10

SUMMARY 18

QUESTIONS 19

FURTHER READING 19

Three structures are the essence of life: cells, chromosomes, and nucleic acids. Cells receive basic sets of instructions from DNA molecules that must also be transmitted to successive generations. And DNA molecules work in the context of larger structures: chromosomes.

Many organisms consist of single cells that can multiply quickly. They are genetically relatively stable, but through changes in their DNA they can adapt rapidly to changes in environmental conditions. Others, including ourselves, animals, plants, and some types of fungi, are multicellular.

Multicellularity offers specialization and complexity: individual cells can be assigned different functions, becoming muscle cells, neurons, or lymphocytes, for example. All the different cells in an individual arise originally from a single cell, and so all nucleated cells carry the same DNA sequences. During development, however, the DNA structure within chromosomes is changed to allow specific changes in gene expression that determine a cell's identity, whether it be a muscle cell or a neuron, for example.

Growth during development and tissue maintenance requires cell division. When a cell divides to produce daughter cells, our chromosomes and the underlying DNA sequences must undergo coordinated duplication and then be carefully segregated to the daughter cells.

Some of our cells can carry our DNA to the next generation. When that happens, chromosomes swap segments and DNA molecules undergo significant changes that make us different from our parents and from other individuals.

DOI: 10.1201/b22853-1

Scanned with CamScanner

1.1 THE STRUCTURE AND FUNCTION OF NUCLEIC ACIDS

General concepts: the genetic material, genomes, and genes

Nucleic acids provide the *genetic material* of cells and viruses. They carry the instructions that enable cells to function in the way that they do and to divide, allowing the growth and reproduction of living organisms. Nucleic acids also control how viruses function and replicate. As we describe later, viruses can be highly efficient at inserting genes into human cells, and modified viruses are widely used in gene therapy.

Nucleic acids are susceptible to small changes in their structure (**mutations**). Occasionally, that can change the instructions that a nucleic acid gives out. The resulting genetic variation, plus mechanisms for shuffling the genetic material from one generation to the next, explains why individual organisms of the same species are nevertheless different from each other. And genetic variation is the substrate that evolutionary forces work on to produce different species. (But note that the different types of cell in a single multicellular organism cannot be explained by genetic variation—the cells each contain the same DNA and the differences in cell types must arise instead by **epigenetic** mechanisms.)

In all cells the genetic material consists of double-stranded DNA in the form of a double helix. (Viruses are different. Depending on the type of virus, the genetic material may be double-stranded DNA, single-stranded DNA, double-stranded RNA, or single-stranded RNA.) As we describe below, DNA and RNA are highly related nucleic acids. RNA is functionally more versatile than DNA (it is capable of self-replication and individual RNA sequences can also serve as templates to make a protein, or act as regulators of gene expression). RNA is widely believed to have developed at a very early stage in evolution. Subsequently, DNA evolved; being chemically much more stable than RNA, it was more suited to being the store of genetic information in cells.

Genome is the collective term for all the *different* DNA molecules within a cell or organism. In prokaryotes—simple unicellular organisms, such as bacteria, that lack organelles—the genome usually consists of just one type of circular double-stranded DNA molecule that can be quite large and has a small amount of protein attached to it. A very large DNA-protein complex such as this is traditionally described as a **chromosome**.

Eukaryotic cells are more complex and more compartmentalized (containing multiple organelles that serve different functions), and they have multiple different DNA molecules. As we will see below, for example, the cells of a man have 25 different DNA molecules but a woman's cells have a genome made up of 24 types of DNA molecule.

In our cells—and in those of all animals and fungi—the genome is partitioned between the nucleus and the mitochondria. Most of the DNA is found in the nucleus, existing as extremely long linear DNA molecules complexed with a variety of different proteins and some types of RNA to form highly organized chromosomes. However, in mitochondria there is just one type of small circular DNA molecule that is largely devoid of protein. (In plant cells, chloroplasts also have their own type of small circular DNA molecule.)

Genes are the DNA segments that carry the genetic information to make proteins or functional noncoding RNA molecules within cells. The great bulk of the genes in a eukaryotic cell are found in the chromosomes of the nucleus; just a few genes are found in the small mitochondrial or chloroplast DNA molecules.

The underlying chemistry of nucleic acids

Each nucleic acid strand is a polymer, a long chain containing many sequential copies of a simple repeating unit, a **nucleotide**. Each nucleotide in turn consists of a sugar molecule, to which is attached a nitrogenous base and a phosphate group.

Scanned with CamScanner

In DNA the sugar is deoxyribose, which has five carbon atoms that are labeled 1′ (one *prime)* to 5′. It is very closely related to ribose, the sugar molecule found in RNA—the only difference is that a hydroxyl (-OH) group at carbon 2′ of ribose is replaced by a hydrogen atom in deoxyribose (Figure 1.1).

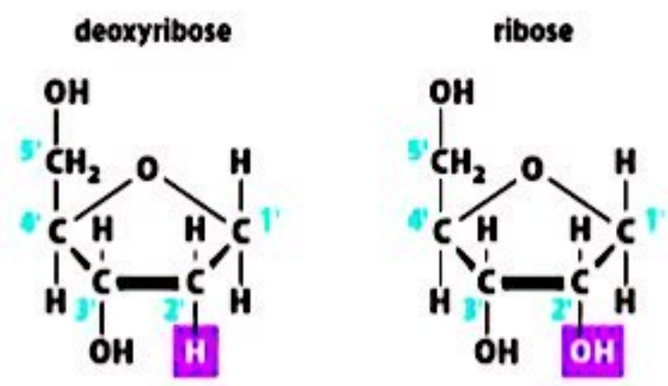

Figure 1.1 Structure of deoxyribose (left) and ribose (right). The f ve carbon atoms are numbered 1′ (one *prime)* to 5′ (f ve *prime)*. The magenta shading is meant to signify the only structural dif erence between deoxyribose (the sugar found in DNA) and ribose (the sugar found in RNA): ribose has a hydroxyl (-OH) group in place of the highlighted hydrogen atom attached to carbon 2′ of deoyribose. The more precise name for deoxyribose is therefore 2′-deoxyribose.

Individual nucleotides are joined to their neighbors by a negatively charged phosphate group that links the sugar components of the neighboring nucleotides. As a result, nucleic acids are polyanions, and have a *sugar-phosphate backbone* with bases bonded to the sugars. As explained in Box 1.1, the sugar-phosphate backbone of each nucleic acid strand is asymmetric and the ends of each strand are asymmetric, giving direction to each strand.

BOX 1.1 5′ AND 3′ ENDS, AND STRAND ASYMMETRY OF NUCLEIC ACIDS

In a nucleic acid strand each phosphate group links carbon atom 3′ from the sugar on one nucleotide to a carbon 5′ on the sugar of a neighboring nucleotide. Internal nucleotides will therefore be linked through both carbon 5′ and carbon 3′ of the sugar to the neighboring nucleotides on opposing sides. However, the nucleotides at the extreme ends of a DNA or RNA strand will have different functional groups. At one end, the **5′ end**, the nucleotide has a terminal sugar with a carbon 5′ that is not linked to another nucleotide and is capped by a phosphate group; at the other end, the **3′ end**, the terminal nucleotide has a sugar with a carbon 3′ that is capped by a hydroxyl group (Figure 1).

The resulting asymmetry between the two ends of a nucleic acid give it a direction. That is important in packing a nucleic acid because when two single nucleic acid strands pair up to make a stable duplex, they must be *anti-parallel*: the 5′ → 3′ direction of one strand must be opposite to that of its partner strand. And direction is important for synthesis of a nucleic acid: a growing nucleic acid strand always extends in a 5′ → 3′ direction.

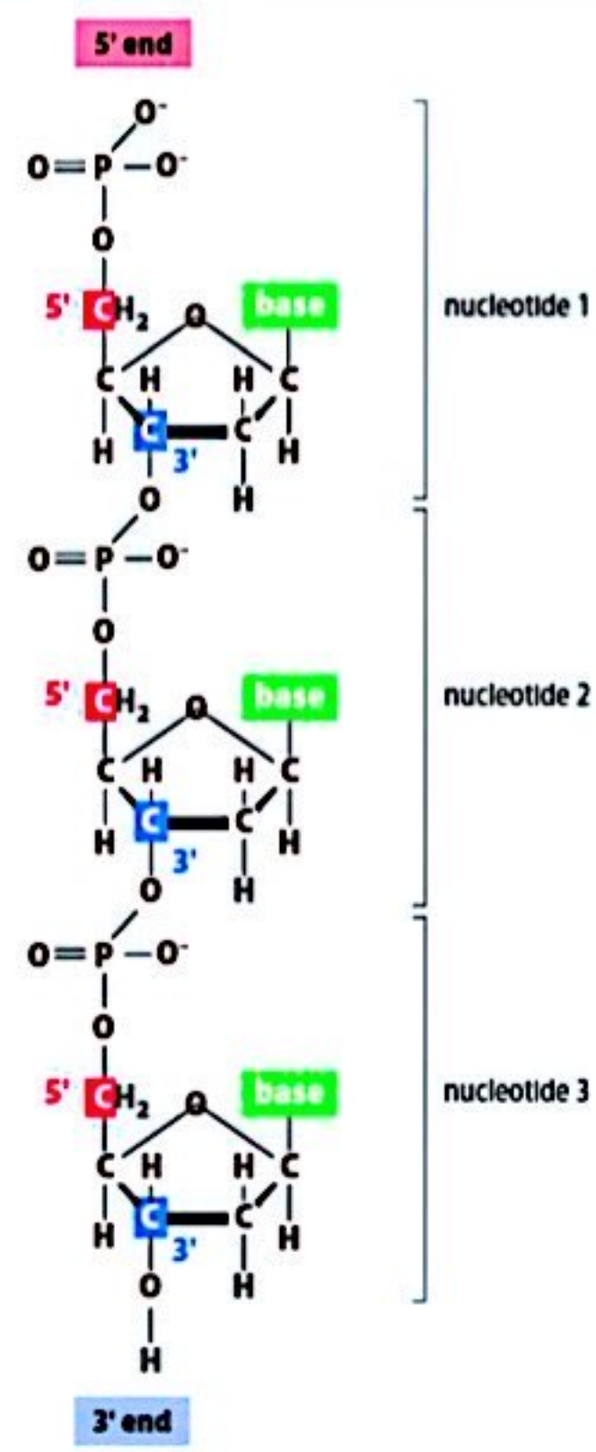

Figure 1 Repeating structure and asymmetric 5′ and 3′ ends in nucleic acids.

Scanned with CamScanner

Unlike the sugar molecules, the nitrogenous bases come in four different types, and it is the sequence of different bases that identifies the nucleic acid and its function. Two of the bases have a single ring based on carbon and nitrogen atoms (a **pyrimidine**) and two have a double ring structure (a **purine**). In DNA the two purines are adenine (A) and guanine (G), and the two pyrimidines are cytosine (C) and thymine (T). The bases of RNA are very similar; the only difference is that in place of thymine there is a very closely related base, uracil (U) (Figure 1.2).

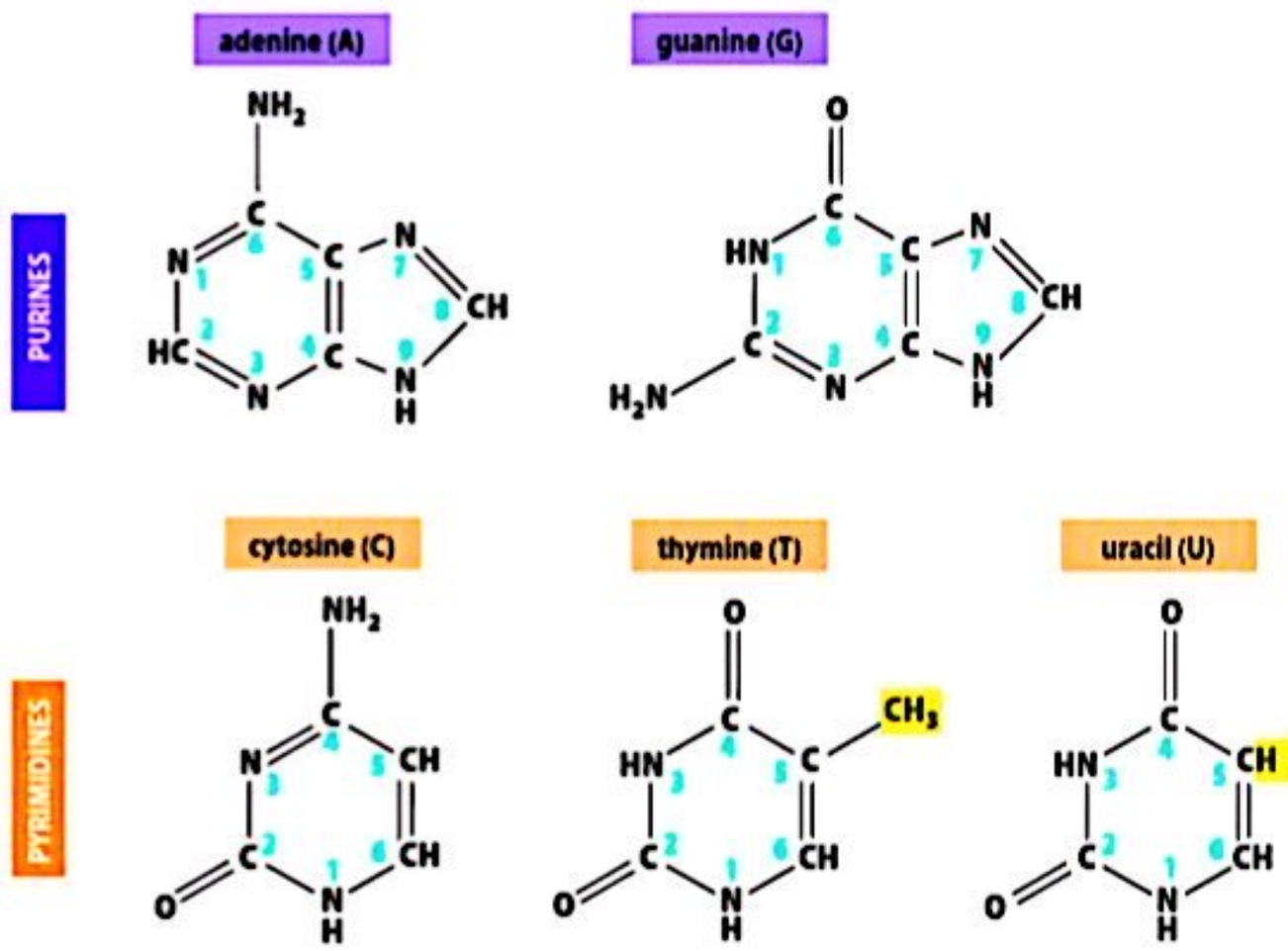

Figure 1.2 Structure of the bases found in nucleic acids. Adenine and guanine are purines with two interlocking rings based on nitrogen and carbon atoms (numbered 1 to 9 as shown). Cytosine and thymine are pyrimidines with a single ring. Adenine, cytosine, and guanine are found in both DNA and RNA, but the fourth base is thymine in DNA and uracil in RNA (they are closely related bases—carbon atom 5 in thymine has an attached methyl group, but in uracil the methyl group is replaced by a hydrogen atom).

Base pairing and the double helix

Cellular DNA exists in a double-stranded (or *duplex*) form, in which the two very long single DNA strands are wrapped round each other. In the resulting double helix each base on one DNA strand is noncovalently linked (by hydrogen bonding) to an opposing base on the opposite DNA strand, forming a **base pair**. However, the two DNA strands fit together correctly only if opposite every A on one strand is a T on the other strand, and opposite every G is a C. (Only two types of base pairs are normally tolerated in double-stranded DNA: A–T and G–C base pairs.) G–C base pairs, which are held together by three hydrogen bonds, are stronger than A–T base pairs, which are held together by two base pairs; see Figure 1.3.

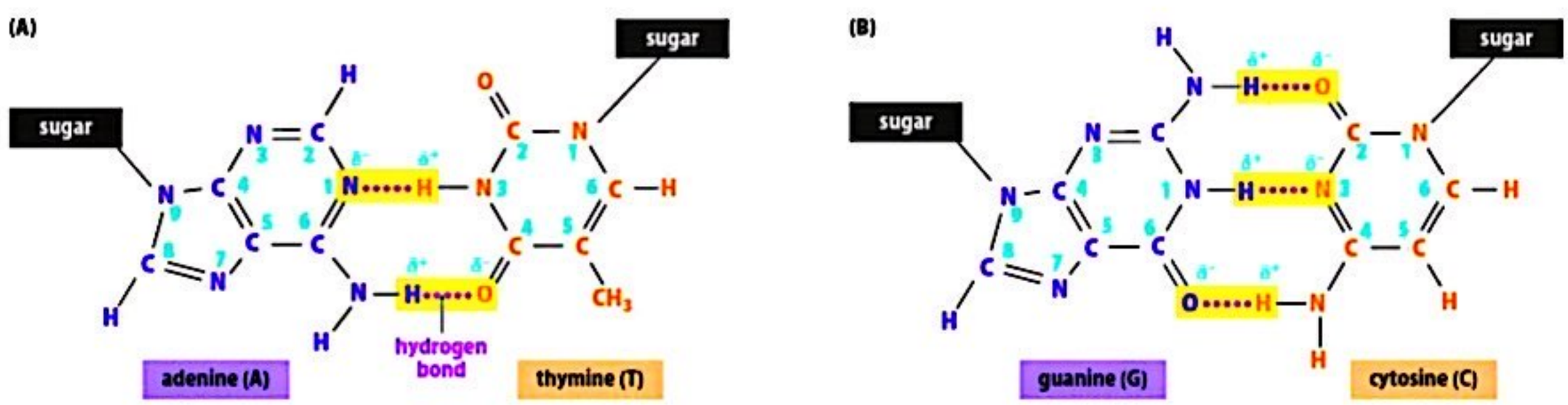

Figure 1.3 Structure of base pairs. In the A–T base pair shown in (A), the adenine is connected to the thymine by two hydrogen bonds. In the G–C base pair shown in (B), three hydrogen bonds link the guanine to the cytosine; a G–C base pair is therefore stronger than an A–T base pair. δ^+ and δ^- indicate fractional positive charges and fractional negative charges.

Scanned with CamScanner

There is one additional restriction on how two single-stranded nucleic acids form a double-stranded nucleic acid. In addition to a sufficient degree of base pairing, for a duplex to form, the two single strands must be anti-parallel; that is, the 5′ → 3′ direction of one strand is the opposite of the 5′ → 3′ direction of the other strand.

Two single nucleic acid strands that can form a double helix with perfect base matching (according to the base pairing rules given above) are said to have **complementary sequences**. As a result of base pairing rules, the sequence of one DNA strand in a double helix can immediately be used to predict the base sequence of the complementary strand (Box 1.2). Note that base pairing can also occur in RNA; when an RNA strand participates in base pairing, the base pairing rules are more relaxed (see Box 1.2).

DNA replication and DNA polymerases

Base pairing rules also explain the mechanism of DNA replication. In preparation for new DNA synthesis before cell division, each DNA double helix must be unwound using a helicase. During the unwinding process the two individual single DNA strands become available as templates for making complementary DNA strands that are synthesized in the 5′ → 3′ direction (Figure 1.4).

BOX 1.2 BASE PAIRING PREVALENCE, SEQUENCE COMPLEMENTARITY, AND SEQUENCE NOTATION FOR NUCLEIC ACIDS

THE PREVALENCE OF BASE PAIRING

The DNA of cells—and of viruses that have a double-stranded DNA genome—occurs naturally as double helices in which base pairing is restricted to A–T and C–G base pairs.

Double-stranded RNA also occurs naturally in the genomes of some kinds of RNA viruses. Although cellular RNA is often single-stranded, it can also participate in base pairing in different ways. Many single-stranded RNAs have sequences that allow intramolecular base pairing—the RNA bends back upon itself to form local double-stranded regions for structural stability and/or for functional reasons. Different RNA molecules can also transiently base pair with each other over short to moderately long regions, allowing functionally important interactions (such as base pairing between messenger RNA and transfer RNA during translation, for example; see Section 2.1). G–U base pairs are allowed in RNA–RNA base pairing, in addition to the standard A–U and C–G base pairs.

RNA–DNA hybrids also form transiently in different circumstances. They occur when a DNA strand is transcribed to give an RNA copy, for example, and when an RNA is reverse transcribed to give a DNA copy.

SEQUENCE COMPLEMENTARITY

Double-helical DNA within cells shows perfect base matching over extremely long distances, and the two DNA strands within a double helix are said to exhibit **base complementarity** and to have *complementary sequences*. Because of the strict base pairing rules, knowing the base sequence of just one DNA strand is sufficient to immediately predict the sequence of the complementary strand, as illustrated below.

SEQUENCE NOTATION

Because the base sequence of a nucleic acid governs its biological properties it is customary to define a nucleic acid by its base sequence, which is always written in the 5′ → 3′ direction. While a single-stranded oligonucleotide sequence might be written accurately as 5′ p-C-p-G-p-A-p-C-p-C-p-A-p-T-OH 3′, where p = phosphate, it is simpler to write it just as CGACCAT.

For a double-stranded DNA the sequence of just one of the two strands is needed (the sequence of the complementary strand can immediately be predicted by the base pairing rules given above). If a given DNA strand has the sequence CGACCAT, the sequence of the complementary strand can easily be predicted to be ATGGTCG (in the 5′ → 3′ direction as shown below, where A–T base pairs are shown in green and C–G base pairs in blue).

Given DNA strand 5′ CGACCAT 3′

→ Complementary strand 3′ GCTGGTA 5′

Scanned with CamScanner

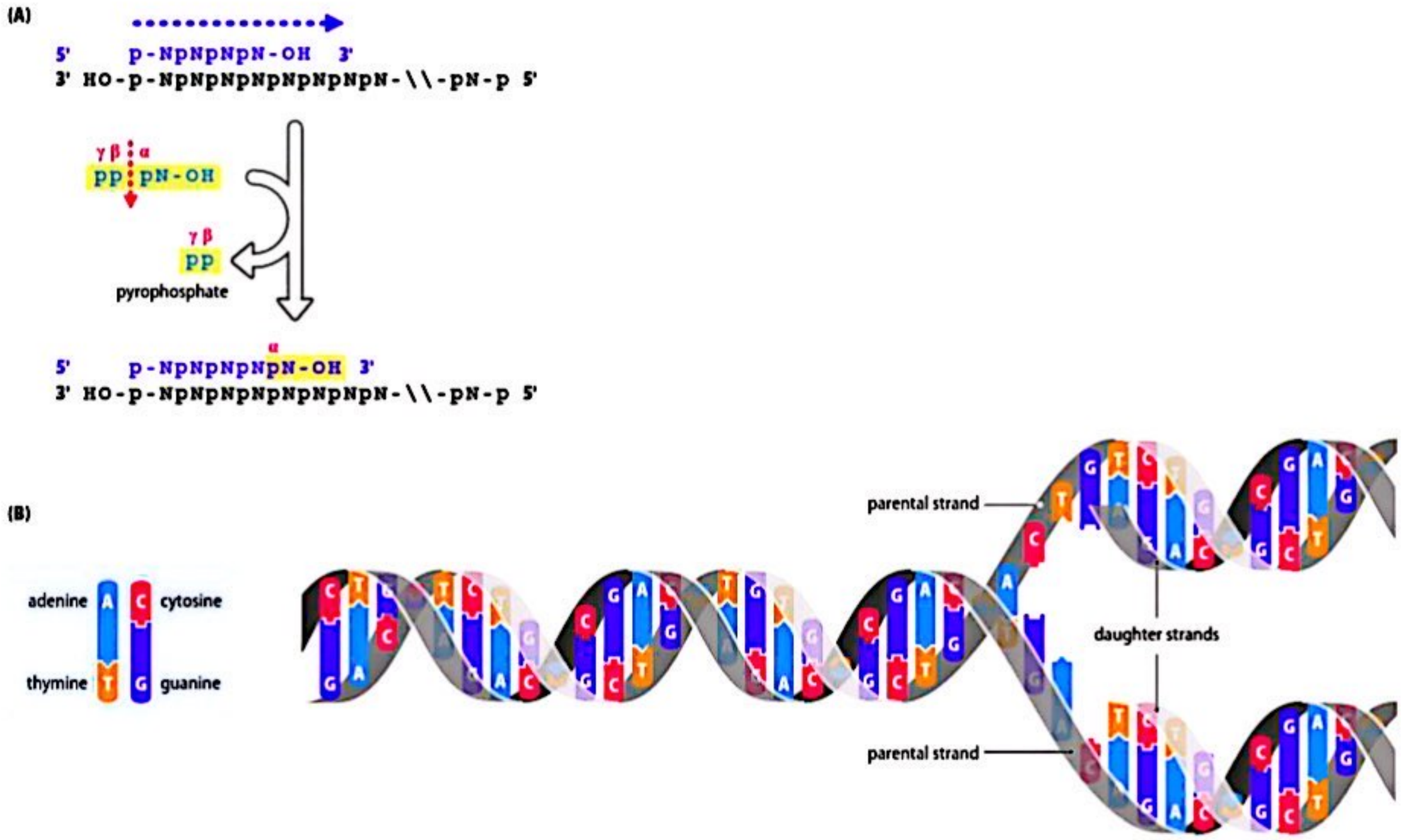

Figure 1.4 DNA synthesis and replication. (A) DNA synthesis. Using a pre-existing DNA strand (black) as a template, a new DNA strand (purple) is synthesized in a 5′ → 3′ direction (dashed arrow) using a DNA polymerase to insert successive dNMPs obtained by cleaving the two external phosphates (α and β), to give pyrophosphate residue (which is discarded). (B) DNA replication. The parental DNA duplex consists of two complementary DNA strands that unwind to serve as templates for the synthesis of new complementary DNA strands (daughter strands). Each completed daughter DNA duplex contains one of the two parental DNA strands plus one newly synthesized DNA strand and is structurally identical to the original parental DNA duplex.

DNA replication therefore uses one double helix to make two double helices, each containing one strand from the parental double helix and one newly synthesized strand (semi-conservative DNA replication). Because DNA synthesis occurs only in the 5′ → 3′ direction, one new strand (the *leading strand*) can be synthesized continuously; the other strand (the *lagging strand*) needs to be synthesized in pieces, known as Okazaki fragments (Figure 1.5).

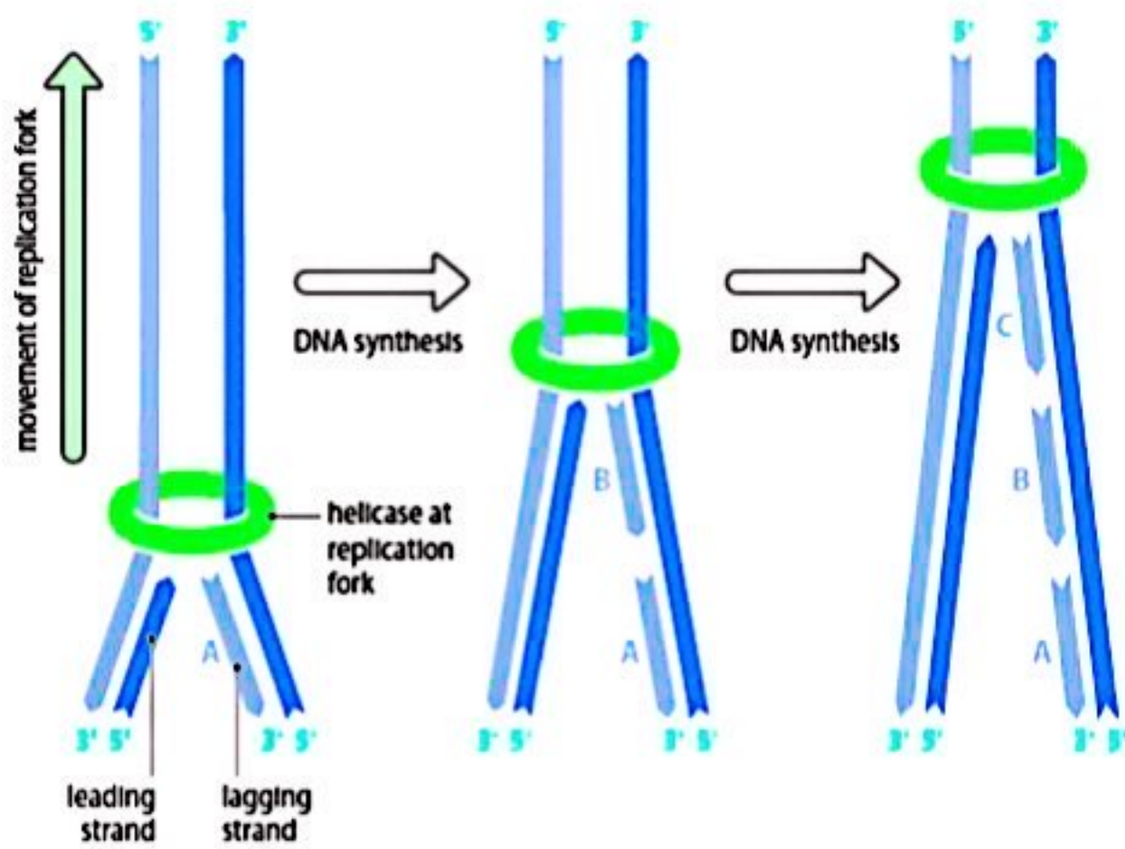

Figure 1.5 Semi-discontinuous DNA replication. The enzyme DNA helicase opens up a **replication fork**, where synthesis of new daughter DNA strands can begin. The overall direction of movement of the replication fork matches that of the continuous 5′ → 3′ synthesis of one daughter DNA strand, the *leading strand*. Replication is semi-discontinuous because the *lagging strand*, which is synthesized in the opposite direction, is built up in pieces (Okazaki fragments, shown here as fragments A, B, and C) that will later be stitched together by a DNA ligase.

Scanned with CamScanner

Mammalian cells have very many kinds of DNA-dependent DNA polymerases that serve a variety of roles, including DNA replication initiation, synthesis of the leading and lagging strands, and also, as described in Section 4.2, multiple roles in DNA repair. Our cells also contain specialized DNA polymerases that use RNA as a template to synthesize a complementary DNA; see Table 1.1.

TABLE 1.1 CLASSICAL DNA-DEPENDENT AND RNA-DEPENDENT DNA POLYMERASES OF MAMMALIAN CELLS

DNA polymerases	Roles
Classical DNA-dependent DNA polymerases	**Standard DNA replication and/or DNA repair**
α (alpha)	initiates DNA synthesis (at replication origins, and also when priming the synthesis of Okazaki fragments on the lagging strand)
δ (delta) and ε (epsilon) β (beta)	major nuclear DNA polymerases and multiple roles in DNA repair
γ (gamma)	base excision repair (repair of deleted bases and simply modifed bases) dedica ted to mitochondrial DNA synthesis and mitochondrial DNA repair
RNA-dependent DNA polymerases	**Genome evolution and telomere function**
Retrosposon reverse transcriptase	occasionally converts mRNA and other RNA into complementary DNA, which can integrate elsewhere into the genome; can occasionally give rise to new genes and new exons, and so on.
TERT (telomerase reverse transcriptase)	replicates DNA at ends of linear chromosomes, using an RNA template

Note: The classical DNA-dependent DNA polymerases are high-fdelit y polymerases—they insert the correct base with high accuracy; however, our cells also have many non-classical DNA-dependent DNA polymerases that exhibit comparatively low fdelit y of DNA replication. We will consider the non-classical DNA polymerases in Chapter 4, because of their roles in certain types of DNA repair and in maximizing the variability of immunoglobulins and T-cell receptors.

Genes, transcription, and the central dogma of molecular biology

As a repository of genetic information, DNA must be stably *transmitted* from mother cell to daughter cells, and from individuals to their progeny; DNA replication provides the necessary mechanism. But within the context of individual cells, the genetic information must also be *interpreted* to dictate how cells work. **Genes** are discrete segments of the DNA whose sequences are selected for this purpose, and gene expression is the mechanism whereby genes are used to direct the synthesis of two kinds of product: RNA and proteins.

The first step of gene expression is to use one of the two DNA strands as a template for synthesizing an RNA copy whose sequence is complementary to the selected template DNA strand. This process is called *transcription*, and the initial RNA copy is known as the primary transcript (Figure 1.6). Subsequently, the primary transcript undergoes different processing steps, eventually giving a mature RNA that belongs to one of two broad RNA classes:

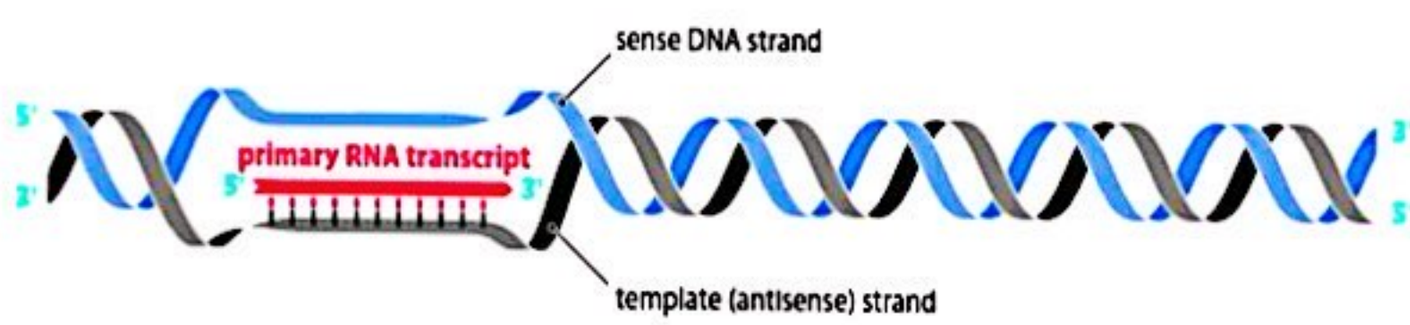

Figure 1.6 Transcription. Transcription results in the synthesis of an RNA transcript in the 5′ → 3′ direction. The nucleotide sequence of the primary RNA transcript is complementary to that of the *template strand* and so is identical to that of the *sense strand*, except that U replaces T. Note: for simplif cation, the diagram does not show the coiling of the RNA transcript around the template DNA strand to form a double helix.

Scanned with CamScanner

- *Coding RNA*. RNAs in this class contain sequences that direct the synthesis of polypeptides (the major component of proteins) in a process called translation. This type of RNA has traditionally been called a messenger RNA (mRNA) because it carries genetic instructions to be decoded by the protein synthesis machinery.
- *Noncoding RNA*. All other mature functional RNAs fall into this class, and here the RNAs, not proteins, are the functional endpoint of gene expression. Noncoding RNAs have a variety of roles in cells, as described in later chapters.

In all forms of life, genetic information is interpreted in what initially seemed to be one direction only: DNA → RNA → protein, a principle that became known as the central dogma of molecular biology. However, certain DNA polymerases, known as reverse transcriptases, found initially in certain types of viruses, can reverse the flow of genetic information by making a DNA copy of an RNA molecule. The cells of complex organisms also have their reverse transcriptases, as described below. In addition, RNA can sometimes also be used as a template to make a complementary RNA copy. So, although genetic information in cells mostly flows from DNA to RNA to protein, the central dogma is no longer strictly valid.

We will explore gene expression (including protein synthesis) in greater detail in Chapter 2. And in Chapter 6 we will focus on both genetic and epigenetic regulation of gene expression.

1.2 THE STRUCTURE AND FUNCTION OF CHROMOSOMES

In this section we consider general aspects of the structure and function of our chromosomes that are largely shared by the chromosomes of other complex multicellular organisms. We will touch on human chromosomes when we consider aspects of the human genome in Chapter 2, when we first introduce the banded pattern of human chromosomes. In Chapter 7 we consider how disease-causing chromosome abnormalities arise. We describe the methodology and the terminology of human chromosome banding in Box 7.2, and diagnostic chromosome analyses in Chapter 11.

Why we need highly structured chromosomes, and how they are organized

Before replication, each chromosome in the cells of complex multicellular organisms normally contains a single, immensely long DNA double helix. For example, an average-sized human chromosome contains a single DNA double helix that is about 4.8 cm long with 140 million nucleotides on each strand; that is, 140 million base pairs (140 megabases (Mb)) of DNA.

To appreciate the difficulty in dealing with molecules this long in a cell only about 10 μm across, imagine a model of a human cell 1 meter across (a 10^5-fold increase in diameter). Now imagine the problem of fitting into this 1-meter-wide cell 46 DNA double helices that when scaled up by the same factor would each be just 0.2 mm thick but on average 4.8 km (about three miles) long. Then there is the challenge of replicating each of the DNA molecules and arranging for the cell to divide in such a way that the replicated DNA molecules are segregated equally into the two daughter cells. All this must be done in a way that avoids any tangling of the long DNA molecules.

To manage nuclear DNA molecules efficiently and avoid any tangling, they are complexed with various proteins and sometimes noncoding structural RNAs to form **chromatin** that undergoes different levels of coiling and compaction to

Scanned with CamScanner

form chromosomes. In interphase—the stages of the cell cycle other than mitosis (see Section 1.3)—the nuclear DNA molecules are still in a very highly extended form and normally the very long slender interphase chromosomes remain invisible under the light microscope. But even in interphase cells, the 2 nm-thick double helix is subject to at least two levels of coiling. First, the double helix is periodically wound round a specialized complex of positively charged histone proteins to form a 10 nm nucleosome filament. The nucleosome filament is then coiled into a 30 nm chromatin fiber that undergoes looping and is supported by a scaffold of nonhistone proteins (Figure 1.7).

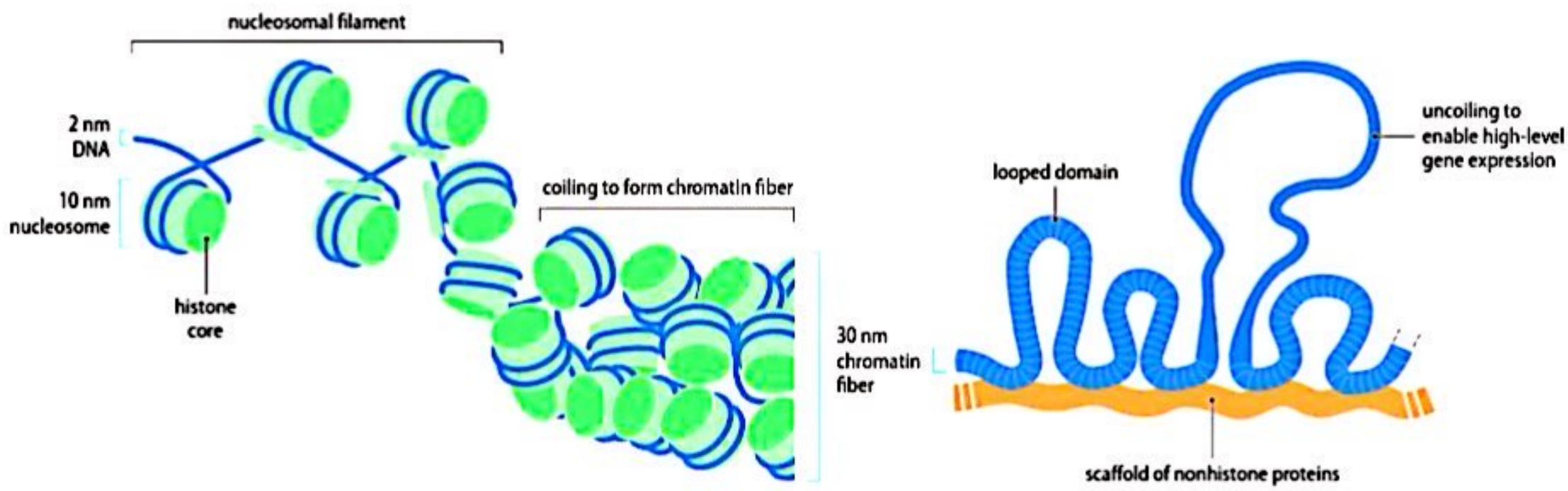

Figure 1.7 From DNA double helix to interphase chromatin. Binding of basic histone proteins causes the 2 nm DNA double helix to undergo coiling, forming f rst a 10 nm f lament studded with nucleosomes that is further coiled to give a 30 nm chromatin f ber. In interphase, the chromatin f ber is organized in looped domains, each containing about 50–200 kilobases of DNA, that are attached to a central scaf old of nonhistone proteins. High levels of gene expression require local uncoiling of the chromatin f ber to give the 10 nm nucleosomal f laments. The diagram does not show structural RNAs that can be important in chromatin. (Adapted from Grunstein M [1992] *SciAm* 267:68–74; PMID 1411455. With permission from Macmillan Publishers Ltd; and Alberts B, Johnson A, Lewis J et al. [2008] Molecular Biology of the Cell, 5th ed. Garland Science.)

During interphase most chromatin exists in an extended state (**euchromatin**) that is dispersed through the nucleus. Euchromatin is not uniform, however—some euchromatic regions are more condensed than others, and genes may or may not be expressed, depending on the cell type and its functional requirements. Some chromatin, however, remains highly condensed throughout the cell cycle and is generally genetically inactive (**heterochromatin**).

As cells prepare to divide, the chromosomes need to be compacted much further to maximize the chances of correct pairing and segregation of chromosomes into daughter cells. Packaging of DNA into **nucleosomes** and then the 30 nm chromatin fiber results in a linear condensation of about 50-fold. During the M (mitosis) phase, higher-order coiling occurs (see Figure 1.7), so that DNA in a human metaphase chromosome is compacted to about 1/10 000 of its stretched-out length. As a result, the short, stubby metaphase chromosomes are readily visible under light microscopes.

Chromosome function: replication origins, centromeres, and telomeres

The DNA within a chromosome contains genes that are expressed according to the needs of a cell. But it also contains specialized sequences that are needed for chromosome function. Three major classes are described below.

Scanned with CamScanner

Centromeres

When a cell divides the chromosomes must be correctly segregated to the two daughter cells. This requires a **centromere**, a region to which a pair of large protein complexes called kinetochores will bind just before the preparation for cell division (Figure 1.8). Centromeres can be seen at metaphase as the primary constriction that separates the short and long chromosome arms. Microtubules attached to each kinetochore are responsible for positioning the chromosomes correctly at metaphase and then pulling the separated chromosomes to opposite poles of the mitotic spindle.

The DNA sequences at centromeres are very different in different organisms. In a mammalian chromosome, the centromeric DNA is a heterochromatic region dominated by highly repetitive DNA sequences that often extend over megabases of DNA.

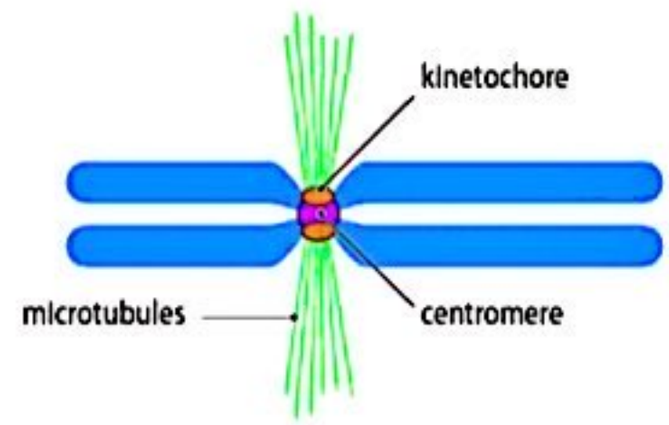

Figure 1.8 Centromere function relies on the assembly of kinetochores and attached microtubules.

Replication origins

For a chromosome to be replicated, it needs one or more replication origins—DNA sequence components to which protein factors bind in preparation for initiating DNA replication. The chromosomes of budding yeast can be replicated using a single very short highly defined DNA sequence, but in the cells of complex organisms, such as mammals, DNA is replicated at multiple initiation sites along each chromosome; the replication origins are quite long and do not have a common base sequence.

Telomeres

Telomeres are specialized structures at the ends of chromosomes that are necessary for the maintenance of chromosome integrity (if a telomere is lost after chromosome breakage, the resulting chromosome end is unstable; it tends to fuse with the ends of other broken chromosomes, or to be involved in recombination events, or to be degraded).

Unlike centromeric DNA, telomeric DNA has been well conserved during evolution. In vertebrates, the DNA of telomeres consists of many tandem (sequential) copies of the sequence TTAGGG to which certain telomeric proteins bind. Most of the telomere DNA is double-stranded with one strand containing TTAGGG repeats (the G-rich strand) and the complementary strand containing CCCTAA repeats (the C-rich strand). However, at its 3′ end, the G-rich strand has an overhang (with about 30 TTAGGG repeats) that folds back and base pairs with the C-rich strand. The resulting T-loop is thought to protect the telomere DNA from natural cellular exonucleases that repair double-strand DNA breaks (Figure 1.9).

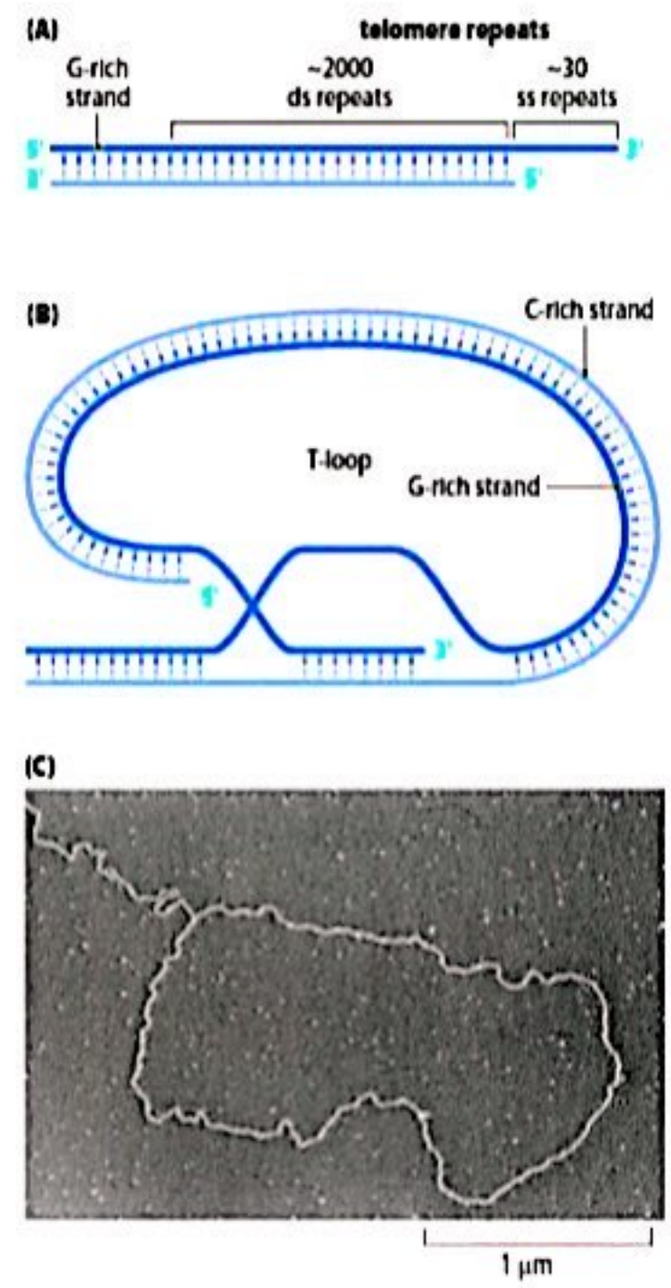

Figure 1.9 Telomere structure and T-loop formation. (A) Human telomere structure. A tandem array of roughly 2000 copies of the double-stranded hexanucleotide (TTAGGG/CCCTAA) repeat followed by a protrusion of about 30 single-stranded TTAGGG repeats. Abbreviations: ss, single-strand; ds, double-strand. (B) T-loop formation. The single-stranded terminus can loop back and invade the double-stranded region by base pairing with the complementary C-rich strand. (C) Electron micrograph showing formation of a roughly 15-kilobase T-loop at the end of an interphase human chromosome. (From Grif th JD et al. [1999] *Cell* 97:503–514; PMID 10338214. With permission from Elsevier.)

1.3 DNA AND CHROMOSOMES IN CELL DIVISION AND THE CELL CYCLE

Differences in DNA copy number between cells

Like other multicellular organisms, we have cells that are structurally and functionally diverse. In each individual the different cell types have the same genetic information, but only a subset of genes is expressed in each cell. What determines the identity of a cell—whether a cell is a B lymphocyte or a hepatocyte, for example—is the pattern of expression of the different genes across the genome.

As well as differences in gene expression, different cells can vary in the number of copies of each DNA molecule. The term *ploidy* describes the number of copies (n) of the basic chromosome set (the collective term for the different chromosomes in a cell) and also describes the copy number of each of the different nuclear DNA molecules.

The DNA content of a single chromosome set is represented as C. Human cells—and the cells of other mammals—are mostly **diploid** (2C), with nuclei containing two copies of each type of chromosome, one paternally inherited

Scanned with CamScanner

and one maternally inherited. Sperm and egg cells are **haploid** cells that contain only one of each kind of chromosome (1C). Human sperm and eggs each have 23 different types of chromosomes and so $n = 23$ in humans.

Some specialized human cells are nulliploid (0C) because they lack a nucleus—examples include erythrocytes, platelets, and terminally differentiated keratinocytes. Others are naturally polyploid (more than 2C). Polyploidy can occur by two mechanisms. The DNA might undergo multiple rounds of replication without cell division, as when the large megakaryocytes in blood are formed (they have from 16 to 64 copies of each chromosome, and the nucleus is large and multilobed). Alternatively, polyploid cells originate by cell fusion to give cells with multiple nuclei, as in the case of muscle fiber cells.

Mitochondrial DNA copy number

The great majority of our cells are diploid and contain two copies of each nuclear DNA molecule. In stark contrast, the number of copies of the mitochondrial DNA (mtDNA) can vary from hundreds to many thousands according to the cell type, and can even vary over time in some cells. The two types of haploid cells show very large differences in mtDNA copy number: a human sperm typically has about 100 mtDNA copies, but a human egg cell usually has about 250 000 mtDNA molecules.

The cell cycle and segregation of replicated chromosomes and DNA molecules

Cells also differ according to whether they actively participate in the cell cycle and undergo successive rounds of cell division. Each time a cell divides, it gives rise to two daughter cells. To keep the number of chromosomes constant there needs to be a tight regulation of chromosome replication and chromosome segregation. Each chromosome needs to be replicated just once to give rise to two daughter chromosomes, which must then segregate equally so that one passes to each daughter cell.

During normal periods of growth there is a need to expand cell number. In the fully grown adult, the majority of cells are terminally differentiated and do not divide, but stem cells and progenitor cells continue to divide to replace cells that have a high turnover, notably blood, skin, sperm, and intestinal epithelial cells.

Each round of the cell cycle involves a phase in which the DNA replicates—S phase (synthesis of DNA)—and a phase where the cell divides—M phase. Note that M phase involves both nuclear division (mitosis) and cell division (cytokinesis). In the intervals between these two phases are two gap phases—G_1 phase (gap between M phase and S phase) and G_2 phase (gap between S phase and M phase)—see Figure 1.10.

Cell division takes up only a brief part of the cell cycle. For actively dividing human cells, a single turn of the cell cycle might take about 24 hours; M phase often occupies about 1 hour. During the short M phase, the chromosomes become extremely highly condensed in preparation for nuclear and cell division. After M phase, cells enter a long growth period called **interphase** ($= G_1 + S + G_2$ phases), during which chromosomes are enormously extended, allowing genes to be expressed.

G_1 is the long-term end state of terminally differentiated nondividing cells. For dividing cells, the cells will enter S phase only if they are committed to mitosis; if not, they are induced to leave the cell cycle to enter a resting phase, the G_0 phase (a modified G_1 stage). When conditions become suitable later on, cells may subsequently move from G_0 to re-enter the cell cycle.

Changes in cell chromosome number and DNA content

During the cell cycle, the amount of DNA in a cell and the number of chromosomes change. In the box panels in Figure 1.10 we follow the fate of a single chromosome through M phase and then through S phase. If we were to

Scanned with CamScanner

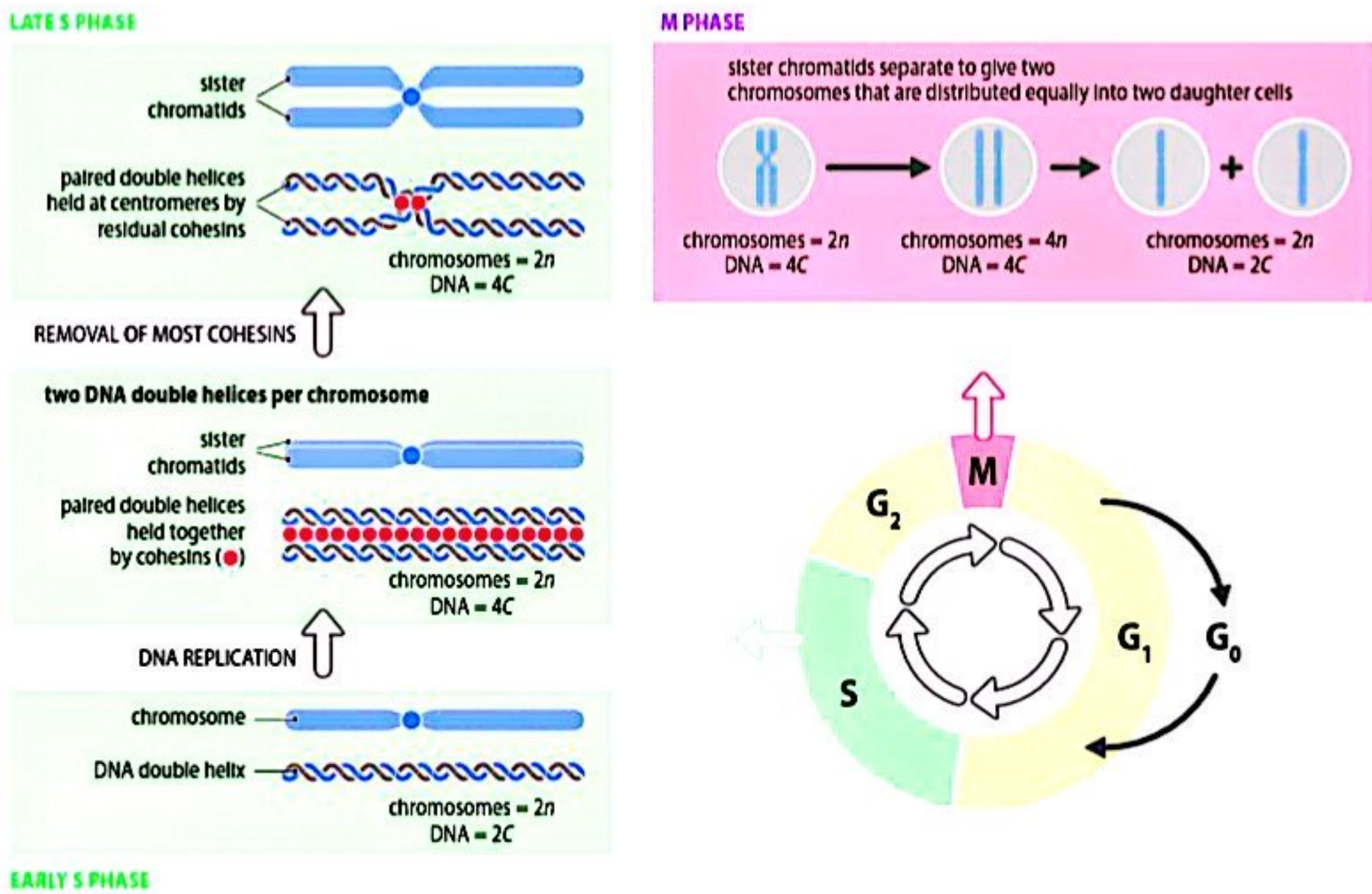

Figure 1.10 Changes in chromosomes and DNA content during the cell cycle. The cell cycle consists of four major phases as shown at the bottom right (in the additional G_0 phase a cell exits from the cell cycle and remains suspended in a stationary phase that resembles G_1 but can subsequently rejoin the cell cycle under certain conditions). In the expanded panels for M and S phases we show for convenience just a single chromosome, and we illustrate in each of the three boxes at left, representing S phase at dif erent stages, how a single chromosome (top) relates to its DNA molecule (bottom). Chromosomes contain one DNA double helix from the end of M phase right through until just before the DNA duplicates in S phase. After duplication, the two double helices are held tightly together along their lengths by binding proteins called cohesins (red circles), and the chromosome now consists of two sister chromatids each having a DNA double helix. The sister chromatids becomes more obvious in late S phase when most of the cohesins are removed except for some at the centromere, which continue to hold the two sister chromatids together. The sister chromatids f nally separate in M phase to form two independent chromosomes that are then segregated into the daughter cells. Note that the S-phase chromosomes in the boxes at left are shown, purely for convenience, in a compact form; in reality they are enormously extended.

consider a diploid human cell this would be one chromosome out of the 46 ($2n$) chromosomes present after daughter cells are first formed. We also show in this figure how a single chromosome (top) relates to its DNA double helix content at different stages in S phase. The progressive changes in the number of chromosomes and the DNA content of cells at different stages of the cell cycle are listed below.

- From the end of the M phase right through until DNA duplication in S phase, each chromosome of a diploid ($2n$) cell contains a single DNA double helix; the total DNA content is therefore 2C.
- After DNA duplication, the total DNA content per cell is 4C, but specialized binding proteins called cohesins hold the duplicated double helices together as **sister chromatids** *within* a single chromosome. The chromosome number remains the same ($2n$), but each chromosome now has double the DNA content of a chromosome in early S phase. In late S phase, most of the cohesins are removed but cohesins at the centromere are retained to keep the sister chromatids together.
- During M phase, the residual cohesins are removed and the duplicated double helices finally separate. That allows sister chromatids to separate to form two daughter chromosomes, giving $4n$ chromosomes. The duplicated chromosomes segregate equally to the two daughter cells so that each will have $2n$ chromosomes and a DNA content of 2C.

Scanned with CamScanner

Figure 1.10 can give the misleading impression that all the interesting action happens in S and M phases. That is quite wrong—a cell spends most of its life in the G_0 or G_1 phases, and that is where the genome does most of its work, issuing the required instructions to make the diverse protein and RNA products needed for cells to function.

Mitochondrial DNA replication and segregation

In advance of cell division, mitochondria increase in mass and mtDNA molecules replicate before being segregated into daughter mitochondria that then need to segregate into daughter cells. Whereas the replication of nuclear DNA molecules is tightly controlled, the replication of mtDNA molecules is not directly linked to the cell cycle.

Replication of mtDNA molecules simply involves increasing the number of DNA copies in the cell, without requiring equal replication of individual mtDNAs. That can mean that some individual mtDNAs might not be replicated and other mtDNA molecules might be replicated several times (Figure 1.11).

Whereas the segregation of nuclear DNA molecules into daughter cells needs to be equal and is tightly controlled, segregation of mtDNA molecules into daughter cells can be unequal. Even if the segregation of mtDNA molecules into daughter mitochondria is equal (as shown in Figure 1.11), the segregation of the mitochondria into daughter cells is thought to be stochastic.

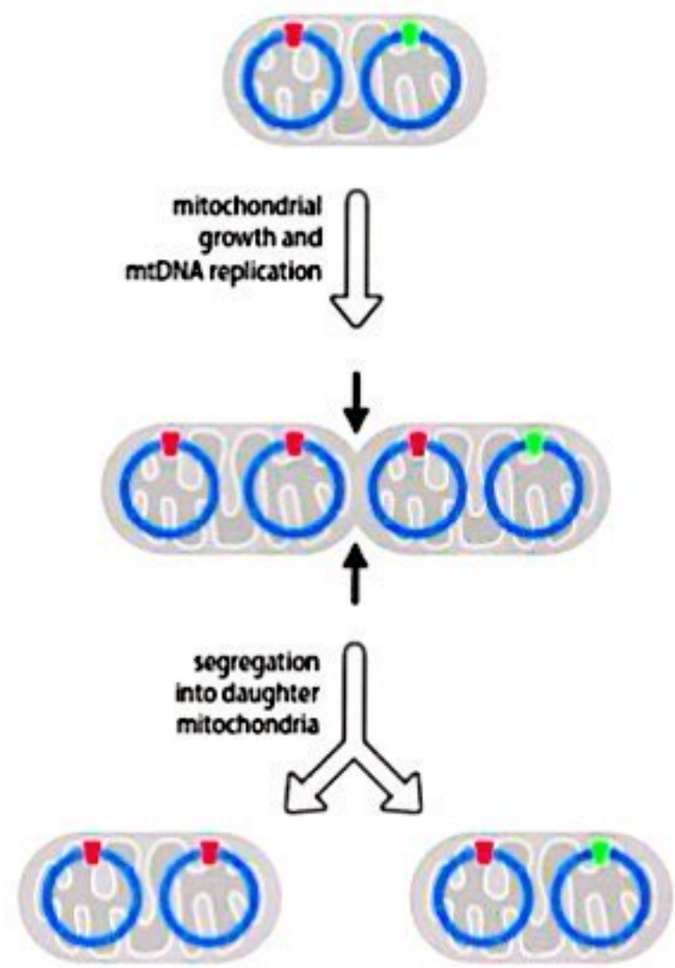

Figure 1.11 Unequal replication of individual mitochondrial DNAs. Unlike in the nucleus, where replication of a chromosomal DNA molecule is tightly controlled and normally produces two copies, mitochondrial DNA (mtDNA) replication is stochastic. When a mitochondrion increases in mass in preparation for cell division, the overall amount of mitochondrial DNA increases in proportion, but individual mtDNAs replicate unequally. In this example, the mtDNA with the green tag fails to replicate and the one with the red tag replicates to give three copies. Variants of mtDNA can arise through mutation so that a person can inherit a mixed population of mtDNAs (heteroplasmy). Unequal replication of pathogenic and nonpathogenic mtDNA variants can have important consequences, as described in Chapter 5.

Mitosis: the usual form of cell division

Most cells divide by a process known as mitosis. In the human life cycle, mitosis is used to generate extra cells that are needed for periods of growth and to replace various types of short-lived cells. Mitosis ensures that a single parent cell gives rise to two daughter cells that are both genetically identical to the parent cell (barring any errors that might have occurred during DNA replication). During a human lifetime, there may be something like 10^{17} mitotic divisions.

The M phase of the cell cycle includes both nuclear division (mitosis, which is divided into the stages of prophase, prometaphase, metaphase, anaphase, and telophase), and also cell division (cytokinesis), which overlaps the final stages of mitosis (Figure 1.12). In preparation for cell division, the previously highly extended duplicated chromosomes contract and condense so that, by the metaphase stage of mitosis, they are readily visible when viewed under the microscope.

The chromosomes of early S phase have one DNA double helix; however, after DNA replication, two identical DNA double helices are produced and held together by cohesins. Later, when the chromosomes undergo compaction in preparation for cell division, the cohesins are removed from all parts of the chromosomes apart from the centromeres. As a result, as early as prometaphase (when the chromosomes are now visible under the light microscope), individual chromosomes can be seen to comprise two **sister chromatids** that remain attached at the centromere (bound by some residual cohesins).

Later, at the start of anaphase, the remaining cohesins are removed and the two sister chromatids can now disengage to become independent chromosomes that will be pulled to opposite poles of the cell and then distributed equally to the daughter cells (see Figure 1.12).

Meiosis: a specialized reductive cell division giving rise to sperm and egg cells

The **germ line** is the collective term for cells that can pass genetic material to the next generation. It includes haploid sperm and egg cells (the **gametes**) and all the diploid precursor cells from which they arise by cell division, going all the way back to the zygote. The nongermline cells are known as **somatic cells**.

Scanned with CamScanner

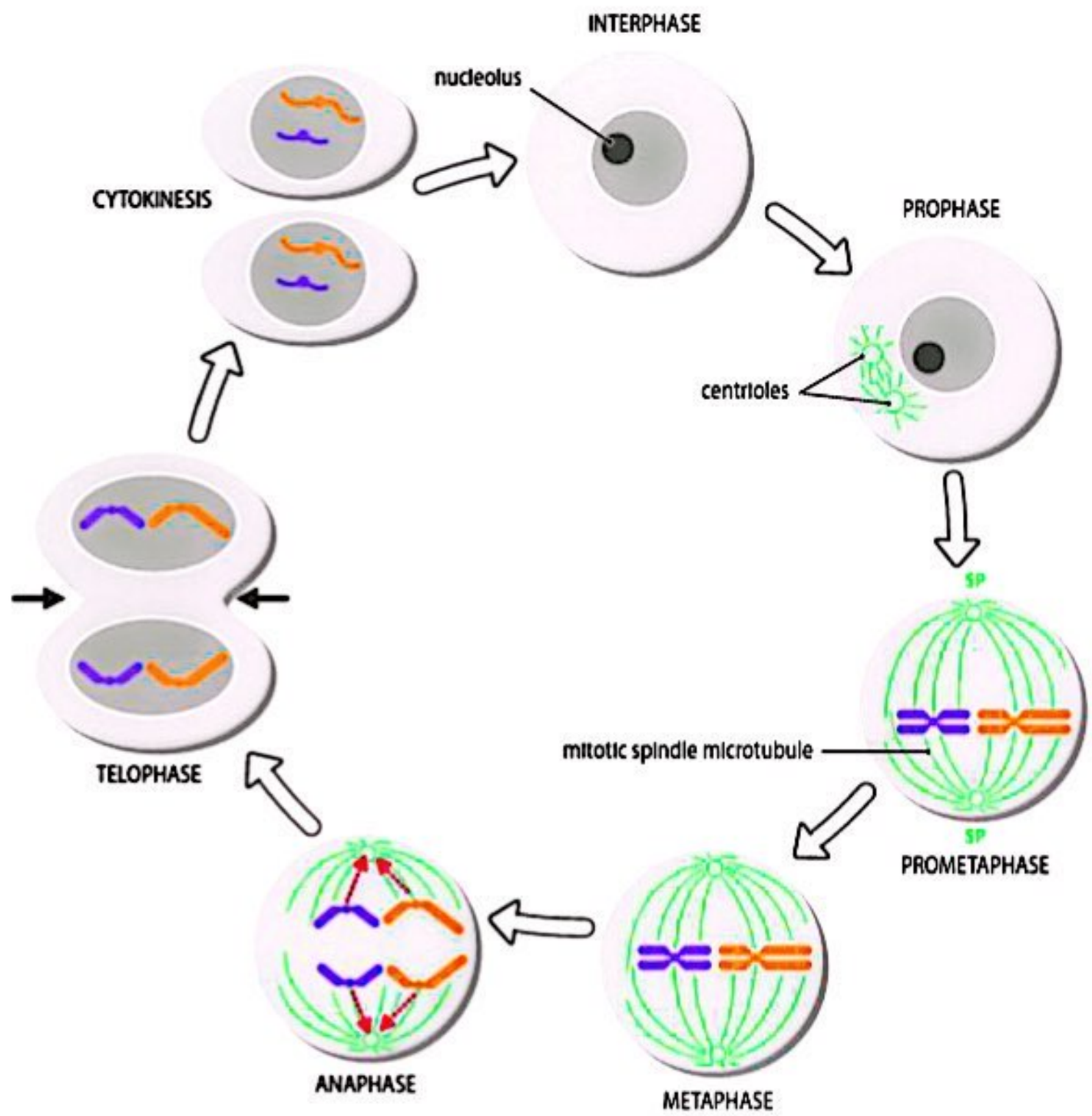

Figure 1.12 Mitosis (nuclear division) and cytokinesis (cell division). Early in prophase, centrioles (short cylindrical structures made up of microtubules and associated proteins) begin to separate and migrate to opposite poles of the cell. They give rise to the spindle poles (SP) from which microtubules will extend to the center of the cell to form the mitotic spindle. In prometaphase, the nuclear envelope breaks down, and the now highly condensed chromosomes become attached at their centromeres to the array of mitotic spindle microtubules. At metaphase, the chromosomes all lie along the middle of the mitotic spindle, still with the sister chromatids bound together (because of residual cohesins at the centromere that hold the duplicated DNA helices together). Removal of the residual cohesins allows the onset of anaphase: the sister chromatids separate and begin to migrate toward opposite poles of the cell (shown by dashed red arrows). The nuclear envelope forms again around the daughter nuclei during telophase, and the chromosomes decondense, completing mitosis. Before the f nal stages of mitosis, and most obviously at telophase, cytokinesis begins. The cell becomes progressively constricted at its middle (shown at telophase by converging horizontal arrows), eventually resulting in full cytokinesis to produce two daughter cells.

In humans, where $n = 23$, each gamete contains one sex chromosome plus 22 nonsex chromosomes (**autosomes**). In eggs the sex chromosome is always an X; in sperm it may be either an X or a Y. After a haploid sperm fertilizes a haploid egg, the resulting diploid **zygote** and almost all of its descendant cells have the chromosome constitution 46,XX (female) or 46,XY (male).

Diploid primordial germ cells migrate into the embryonic gonad and engage in repeated rounds of mitosis, to generate spermatogonia in males and oogonia in females. Further growth and differentiation produce primary spermatocytes in the testis and primary oocytes in the ovary. The diploid spermatocytes and oocytes can then undergo **meiosis**, the cell division process that produces haploid gametes.

Meiosis is a *reductive* division because it involves two successive cell divisions (meiosis I and meiosis II) but only one round of DNA replication (Figures 1.13 and 1.14). As a result, it gives rise to four haploid cells. In males, the two meiotic cell divisions are each symmetric, producing four functionally equivalent spermatozoa. Huge numbers of sperm are produced, and spermatogenesis is a continuous process from puberty onward.

Female meiosis is different: cell division is asymmetric, resulting in unequal division of the cytoplasm. The products of female meiosis I (the first meiotic cell division) are a large secondary oocyte and a small cell, the *polar body,* which is discarded. During meiosis II the secondary oocyte then gives rise to the large mature egg cell and a second polar body (which again is discarded).

In humans, primary oocytes enter meiosis I during fetal development but are then all arrested at prophase until after the onset of puberty. After puberty in females, one primary oocyte completes meiosis with each menstrual cycle. Because ovulation can continue up to the fifth and sometimes sixth decades, this means that meiosis can be arrested for many decades in those primary oocytes that are not used in ovulation until late in life.

Scanned with CamScanner

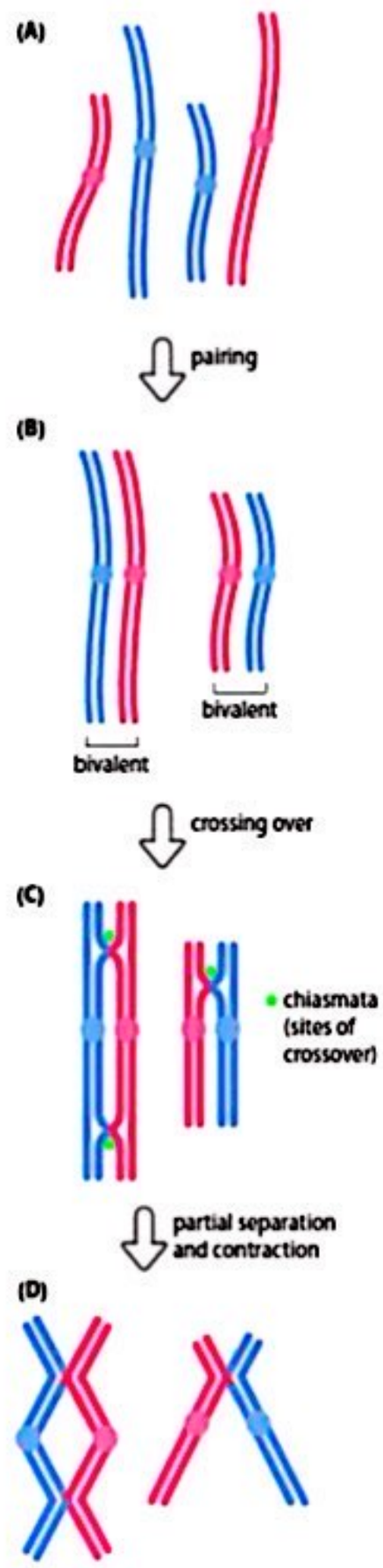

Figure 1.13 Prophase stages in meiosis I. (A) In leptotene, the duplicated chromosomes (each with a pair of sister chromatids) begin to condense but remain unpaired. (B) In zygotene, pairing of maternal and paternal homologous chromosomes (**homologs**) occurs, to form bivalents with four chromatids. (C) In pachytene, recombination (crossing over) occurs through the physical breakage and subsequent rejoining of maternal and paternal chromosome fragments. There are two chiasmata (crossovers) in the bivalent on the left, and one in the bivalent on the right. For simplicity, both chiasmata on the left involve the same two chromatids. In reality, more chiasmata may occur, involving three or even all four chromatids in a bivalent. (D) During diplotene, the homologous chromosomes may separate slightly, except at the chiasmata. A further stage, diakinesis, is marked by contraction of the bivalents and is the transition to metaphase I. In this f gure, only 2 of 23 possible pairs of homologs are illustrated (with the maternal homolog colored pink, and the paternal homolog blue).

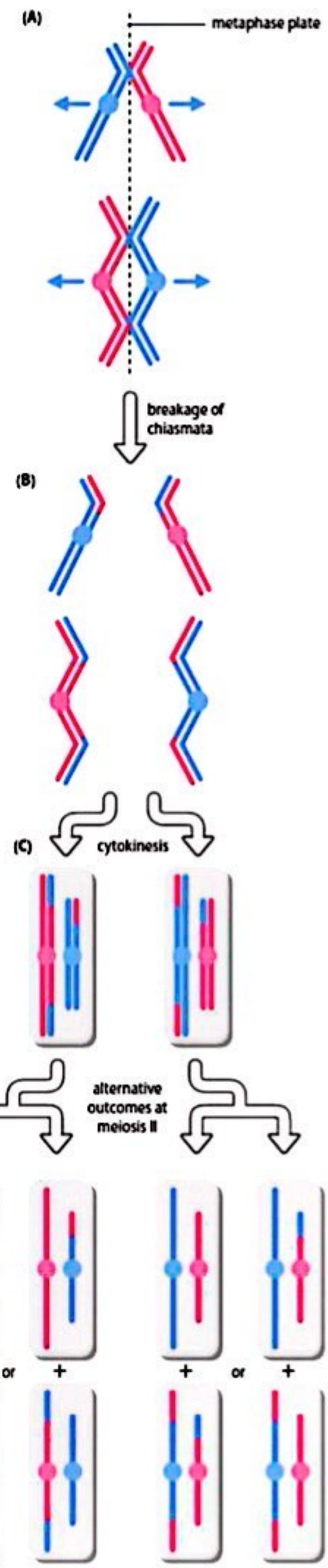

Figure 1.14 Metaphase I to production of gametes. (A) At metaphase I, the bivalents align on the metaphase plate, at the center of the spindle apparatus. Contraction of spindle f bers draws the chromosomes in the direction of the spindle poles (arrows). (B) The transition to anaphase I occurs at the consequent rupture of the chiasmata. (C) Cytokinesis segregates the two chromosome sets, each to a dif erent primary spermatocyte. Note that, after recombination during prophase I (see Figure 1.13C), the chromatids share a single centromere but are no longer identical. (D) Meiosis II in each primary spermatocyte, which does not include DNA replication, generates unique genetic combinations in the haploid secondary spermatocytes. Only 2 of the possible 23 dif erent human chromosomes are depicted, for clarity, so only 2^2 (that is, 4) of the possible 2^{23} (8 388 608) possible combinations are illustrated. Although oogenesis can produce only one functional haploid gamete per meiotic division, the processes by which genetic diversity arises are the same as in spermatogenesis.

Scanned with CamScanner

Pairing of paternal and maternal homologs (synapsis)

Each of our diploid cells contains two copies (**homologs**) of each type of chromosome, one maternal copy and one paternal copy. So, for example, paternal chromosome 1 and maternal chromosome 1 are homologs. The exception, of course, are the X and Y chromosomes in males.

A special feature of meiosis I—that distinguishes it from mitosis and meiosis II—is the pairing *(synapsis)* of paternal and maternal homologs. Then, the maternal and paternal homologs, each with sister chromatids following DNA replication, align along their lengths and become bound together. The resulting *bivalent* has four strands: two paternally inherited sister chromatids and two maternally inherited sister chromatids (see Figures 1.13B–D and 1.14).

The pairing of homologs is required for recombination to occur (as described in the next subsection). It must ultimately be dictated by high levels of DNA sequence identity between the homologs. The high sequence matching between homologs required for pairing does not need to be complete, however: when there is some kind of chromosome abnormality so that the homologs do not completely match, the matching segments usually manage to pair up.

Pairing of maternal and paternal sex chromosomes is straightforward in female meiosis, but in male meiosis there is the challenge of pairing a maternally inherited X chromosome with a paternally inherited Y. The human X chromosome is very much larger than the Y, and their DNA sequences are very different. However, they do have some sequences in common, notably a major *pseudoautosomal region* located close to the short-arm telomeres. The X and Y chromosomes cannot pair up along their lengths, but because they have some sequences in common, they can always pair up along these regions. We will explore this in greater detail in Chapter 5 when we consider pseudoautosomal inheritance.

Recombination

The prophase of meiosis I begins during fetal life and, in human females, can last for decades. During this extended process, parternal and maternal chromatids within each bivalent normally exchange segments of DNA at randomly positioned but matching locations. This process—called **recombination** (or crossover)—involves physical breakage of the DNA in one paternal and one maternal chromatid, and the subsequent joining of maternal and paternal fragments.

Recombined homologs seem to be physically connected at specific points. Each such connection marks the point of crossover and is known as a chiasma (plural chiasmata—see Figure 1.13C). The distribution of chiasmata across chromosomes is nonrandom. The number of chiasmata per meiosis shows significant sex differences, and there are very significant differences between individuals of the same sex (and even between individual meioses from a single individual). In a large recent study of human meiosis, an average of 38 recombinations were detected per female meiosis, while 24 meioses occurred on average in male meiosis but with very significant variation (shown in Figure 8.3 on page 244). In addition to their role in recombination, chiasmata are thought to be essential for correct chromosome segregation during meiosis I.

There are hotspot regions where recombination is more likely to occur. For example, recombination is more common in subtelomeric regions. In the case of X–Y crossover there is an obligate crossover within a short 2.6 Mb pseudoautosomal region located at the tips of the short arms of the X and Y. This region is so called because it is regularly swapped between the X and Y chromosomes and so the inheritance pattern for any DNA variant here is not X-linked or Y-linked but instead resembles autosomal inheritance.

Why each of our gametes is unique

The sole purpose of sex in biology is to produce novel combinations of gene variants, and the instrument for achieving that aim is meiosis. The whole point of

Scanned with CamScanner

meiosis is to produce *genetically unique* gametes by selecting different combinations of DNA sequences on maternal and paternal homologs.

Although a single ejaculate may contain hundreds of millions of sperm, meiosis ensures that no two sperm will be genetically identical. Equally, no two eggs are genetically identical. Each zygote must also be unique because at fertilization a unique sperm combines with a unique egg. However, a unique fertilization event can occasionally give rise to two genetically identical (**monozygotic**) twins if the embryo divides into two at a very early stage in development (monozygotic twins are nevertheless unique individuals—genetics is not everything in life!).

The second division of meiosis is identical in form to mitosis; meiosis I is where the genetic diversity originates, and that involves two mechanisms. First, there is independent assortment of paternal and maternal homologs. After DNA replication, the homologous chromosomes each comprise two sister chromatids, so each bivalent is a four-stranded structure at the metaphase plate. Spindle fibers then pull one complete chromosome (two chromatids) to either pole. In humans, for each of the 23 homologous pairs, the choice of which daughter cell each homolog will enter is independent. This allows 2^{23} or about 8.4×10^6 different possible combinations of parental chromosomes in the gametes that might arise from a single meiotic division (Figure 1.15).

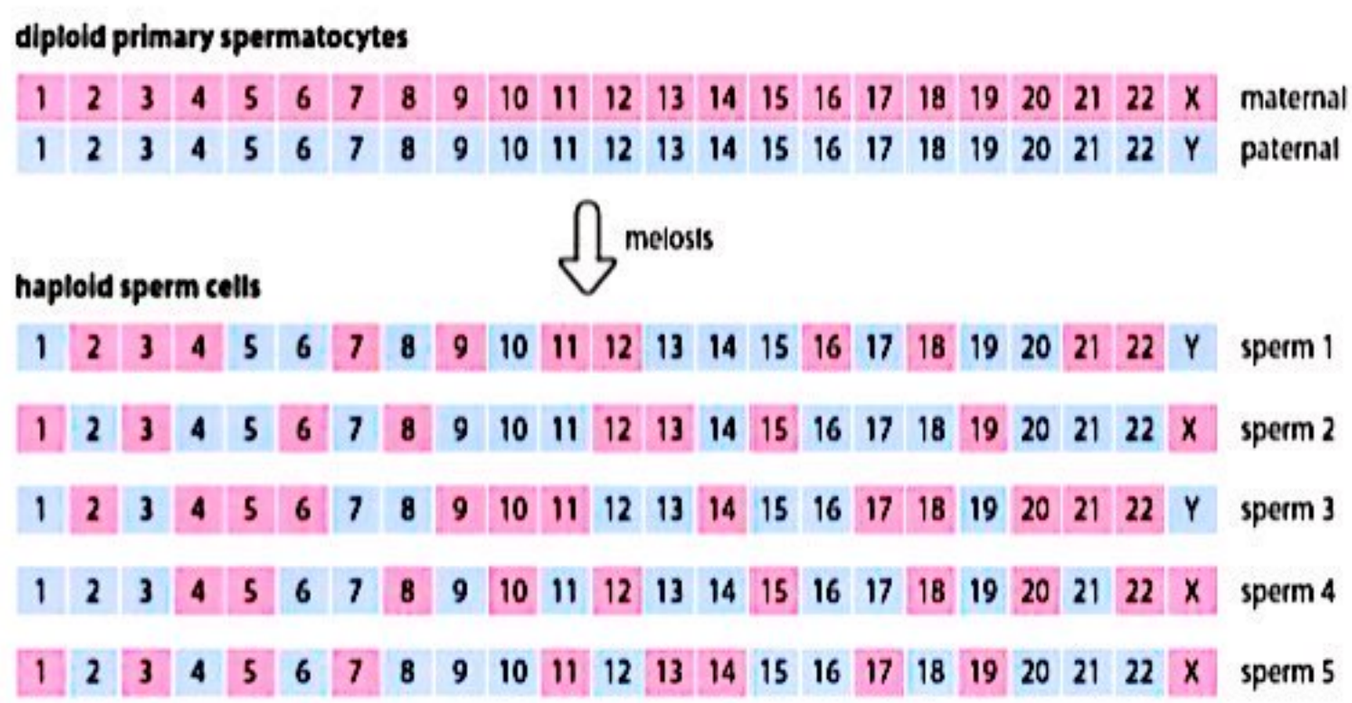

Figure 1.15 Independent assortment of maternal and paternal homologs during meiosis. The f gure shows a random selection of just 5 of the 8 388 608 (2^{23}) theoretically possible combinations of homologs that might occur in haploid human spermatozoa after meiosis in a diploid primary spermatocyte. Maternally derived homologs are represented by pink boxes, and paternally derived homologs by blue boxes. For simplicity, the diagram ignores recombination—but see Figure 1.16.

The second mechanism that contributes to genetic diversity is recombination. Whereas sister chromatids within a bivalent are genetically identical, the paternal and maternal chromatids are not. On average their DNA will differ at roughly 1 in every 1000 nucleotides. Swapping maternal and paternal sequences by recombination will therefore produce an extra level of genetic diversity (Figure 1.16). It raises the number of permutations from the 8.4 million that are possible just from the independent assortment of maternal and paternal homologs alone, to a virtually infinite number.

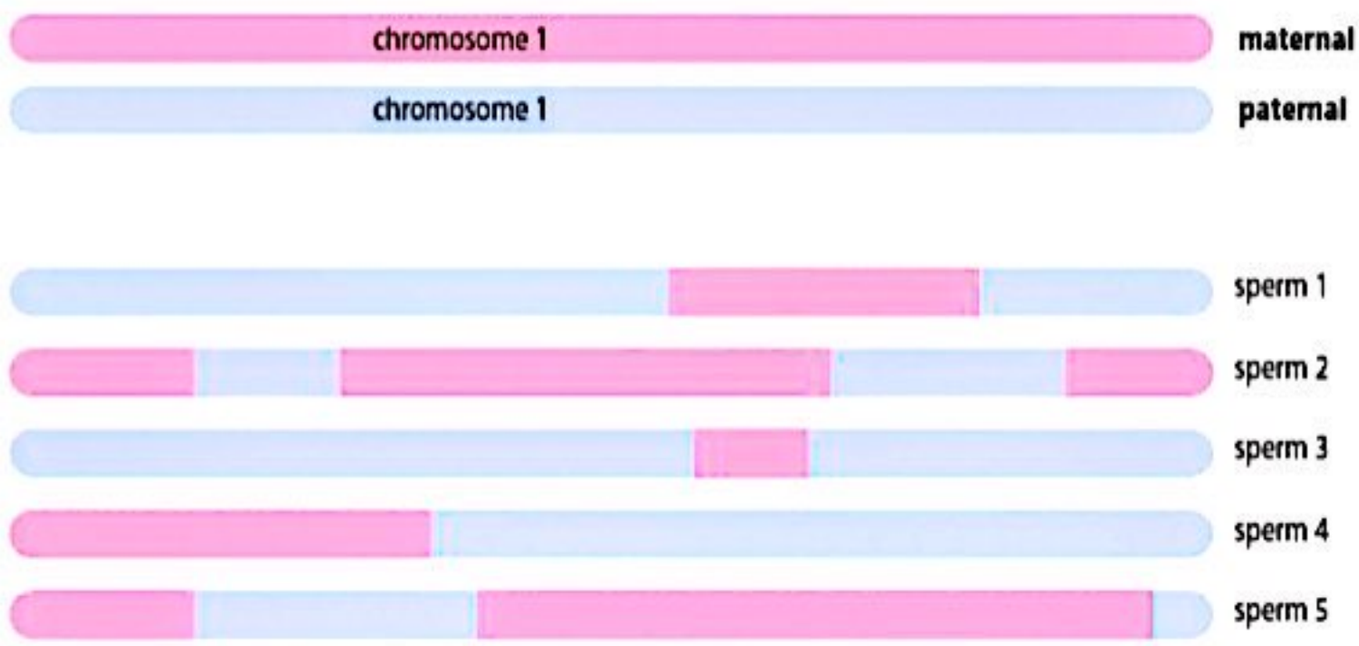

Figure 1.16 Recombination superimposes additional genetic variation at meiosis I. Figure 1.15 illustrates the contribution to genetic variation at meiosis I made by independent assortment of homologs, but for simplicity it ignores the contribution made by recombination. In reality each transmitted chromosome is a mosaic of paternal and maternal DNA sequences, as shown here. See Figure 8.1 on page 243 for a real-life example.

Scanned with CamScanner

SUMMARY

- Nucleic acids are negatively charged long polymers composed of sequential nucleotides that each consist of a sugar, a nitrogenous base, and a phosphate group. They have a sugar-phosphate backbone with bases projecting from the sugars.
- Nucleic acids have four types of bases: adenine (A), cytosine (C), guanine (G), and either thymine (T) in DNA or uracil (U) in RNA. The sequence of bases determines the identity of a nucleic acid and its function.
- RNA normally consists of a single nucleic acid chain, but in cells a DNA molecule has two chains (strands) that form a stable duplex (in the form of a double helix). Duplex formation requires hydrogen bonding between matched bases (base pairs) on the two strands.
- In DNA two types of base pairing exist: A pairs with T, and C pairs with G. According to these rules, the two strands of a DNA double helix are said to have complementary base sequences.
- Base pairing also occurs in RNA and includes G–U base pairs, as well as G–C and A–U base pairs. Two dif erent RNA molecules with partly complementary sequences can associate by forming hydrogen bonds. Intramolecular hydrogen bonding also allows a single RNA chain to form a complex three-dimensional structure.
- DNA carries primary instructions that determine how cells work and how an individual is formed. Def ned segments of DNA called genes are used to make a single-stranded RNA copy that is complementary in sequence to one of the DNA strands (transcription).
- DNA is propagated from one cell to daughter cells by replicating itself. The two strands of the double helix are unwound, and each strand is used to make a new complementary DNA copy. The two new nuclear DNA double helices (each with one parental DNA strand and one new DNA strand) are segregated so that each daughter cell receives one DNA double helix.
- RNA molecules function in cells either as a mature noncoding RNA, or as a messenger RNA with a coding sequence used to make the polypeptide chain of a protein (translation).
- Each nuclear DNA molecule is complexed with different proteins and some noncoding RNAs to form a chromosome that condenses the DNA and protects it.
- Packaging DNA into chromosomes stops the long DNA chains from getting entangled within cells, and by greatly condensing the DNA in preparation for cell division it allows the DNA to be segregated correctly to daughter cells and to of spring.
- Our sperm and egg cells are haploid cells with a set of 23 dif erent chromosomes (each with a single distinctive DNA molecule). There is one sex chromosome (an X chromosome in eggs; either an X or Y in sperm) and 22 dif erent autosomes (nonsex chromosomes).
- Most of our cells are diploid with two copies of the haploid chromosome set, one set inherited from the mother and one from the father. Maternal and paternal copies of the same chromosome are known as homologs.
- There is one type of mitochondrial DNA (mtDNA); it is present in many copies with wide variation in copy number between dif erent cell types. Both the replication of mtDNA and its segregation to daughter cells occur stochastically.
- Cells need to divide as we grow. In fully formed adults, most of our cells are specialized, nondividing cells, but some cells are required to keep on dividing to replace short-lived cells, such as blood, skin, and intestinal epithelial cells.
- Mitosis is the normal form of cell division. Each chromosome (and chromosomal DNA) replicates once and the duplicated chromosomes are segregated equally into the two daughter cells.
- Meiosis is a specialized form of cell division required to produce haploid sperm and egg cells. The chromosomes in a diploid spermatogonium or oogonium replicate once, but there are two successive cell divisions to reduce the number of chromosomes in each cell.
- Each sperm cell produced by a man is unique, as is each egg cell that a woman produces. During the f rst cell division in meiosis, maternal and paternal homologs associate and exchange sequences by recombination. Largely random recombination results in unpredictable new DNA sequence combinations in each sperm and in each egg.

Scanned with CamScanner