

CHAPTER 1

The cell and cell division

Margaret J. Barch¹ and Helen J. Lawce²

¹*(deceased) formerly, Frank F Yen Cytogenetics Laboratory, Weisskopf Child Evaluation Center, University of Louisville, Louisville, KY, USA

²Oregon Health & Science University Knight Diagnostic Laboratory, Portland, OR, USA

1.1 The cell [1,2]

The cell is the basic unit of life – the simplest structure capable of independent existence. The simplest organisms consist of only one cell. Higher organisms are composed of complex colonies of interdependent cells, each colony with a specialized function necessary for the survival of the organism. Cells that have the same general function are often grouped together to form tissues, such as muscle, bone, and connective tissue. Tissues may be combined in larger functional units called organs, such as kidneys, skin, and heart. Organs can in turn be grouped by function into organ systems, such as the respiratory and circulatory systems.

Cells vary greatly in size, but they all must be able to survive and reproduce to be successful organisms. The cell membrane that envelops its contents must be able to control the movement of nutrients into the cell and of ions, molecules, and proteins out of the cell. Energy is converted from food and/or light and is used to synthesize internal components. The information for reproducing cell structures is encoded within its genetic makeup, thus providing the cell with its own self-sufficient capability to reproduce life-supporting needs and to repair genetic damage as needed. When functioning properly, the cell contains all the necessary tools to survive.

1.1.1 Cell membrane

Composition

The cell generally consists of cytoplasm, bounded by a cell membrane, and a nucleus, also enclosed in a membrane. There are exceptions to this model, such as red blood cells that have lost their nuclei during differentiation. The plasma membrane, or cell membrane, defines the boundary of the cell (Figure 1.1) and consists primarily of phospholipids and proteins. The phospholipids form a bimolecular layer, with their hydrophilic ends at the outer surfaces of the membrane and their hydrophobic chains extending into the middle of the membrane. The protein components of the membrane are globular particles distributed through the lipid bilayer; their polar amino acids may be exposed on an outer surface, but nonpolar portions remain in the interior.

Physical barrier

The cell membrane serves as a physical barrier for the cell contents, but it is rather fragile. If one were to tear a hole in this membrane by micromanipulation, the contents would spill out into the surrounding medium. An intact cell can rapidly repair minor membrane damage, but more extensive damage leads to cell death.

* Editors' note: We lost Margaret in the final stages of producing this book. May her spirit shine through and the reader be touched by her love of science, and her passion for passing it on.

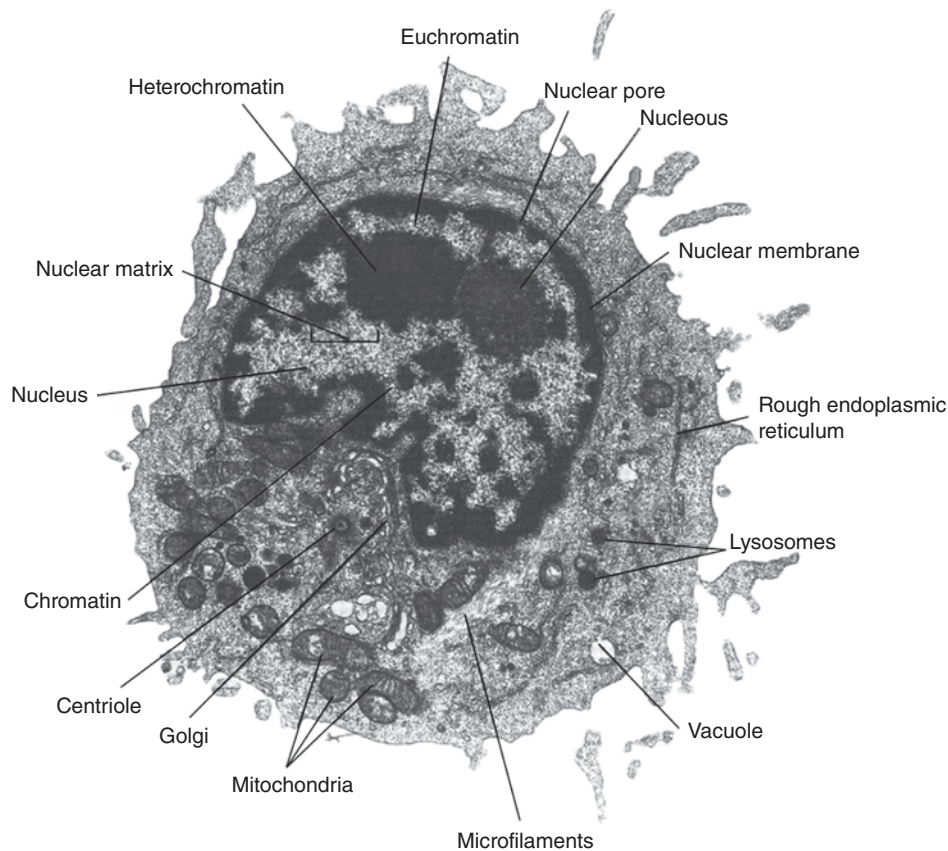


Figure 1.1 An electron micrograph showing the various components of a eukaryotic (human) cell.

Regulatory barrier

The membrane also acts as a regulatory barrier for the entry and exit of molecules and particles. This ability to regulate the passing of substances is called *selective permeability*. Substances can cross the cell membrane by three mechanisms: by free diffusion along a gradient, meaning that substances travel from regions of high concentration to regions of lower concentration; by active transport, which requires energy and moves substances against a concentration gradient; and by enclosure in vesicles that move substances into the cell (endocytosis) or out of the cell (exocytosis). Water can move freely across cell membranes in both directions; it is this property that allows hypotonic solutions (those less concentrated than the inside of the cell) to swell the mitotic cell, thus facilitating chromosome spreading for cytogenetic study.

Glycoprotein functionality

Molecules of glycoprotein (proteins with sugar molecules attached at points along the amino acid chain) exist on the surface of the protein–lipid membrane and sometimes project through it, into the cell. These glycoproteins function in cell adhesion, both to other cells and to culture flask surfaces. Trypsin, a protease (an enzyme that digests proteins), removes these molecules, thereby freeing cells for subculture or harvest. Glycoproteins can be antigenic (e.g., in red cells they determine blood type), and can serve as receptors for viruses, plant agglutinins (e.g., phytohemagglutinin), and hormones. They are further implicated in contact inhibition, a process in which normal cells stop dividing as cultures become confluent. Tumor cells often lose this property and tend to keep growing unchecked in a disorganized fashion when the growth surface is limited. Glycoproteins on the cell surface are also important in cell–cell recognition. If lymphocytes are stripped of their glycoproteins, they no longer accumulate in the lymph nodes.

1.1.2 Cytoplasm

Cytoplasm is the part of the cell within the cell membrane, excluding the nucleus, that consists of water, inorganic ions or molecules, and a variety of organic compounds. In many ways it resembles a colloid, with particles suspended in a continuous gel-like substance called the cytosol. The cytosol, in turn, contains a cytoskeleton of tubules and filaments, dissolved molecules, and water. Among the inorganic molecules are potassium, sodium, magnesium, and calcium. Trace amounts of many heavy metals are also present, as are bicarbonate and phosphate. Tiny granules can also be seen with a light microscope. These granules have been shown to be a series of vacuolar structures, bound by lipoprotein membranes similar to the cell membrane, with some even further differentiated into a complex system of internal membranes.

The large organic molecules (called macromolecules), which give the cytoplasm its colloidal properties, can be grouped into three main classes: proteins, nucleic acids, and polysaccharides. Each class is a polymer built from different subunits (monomers): proteins are made up of amino acid subunits; nucleic acids are polymers of nucleotides; and polysaccharides are built from sugar monomers. Together, the organelles described below and the cytosol make up the cytoplasm.

Proteins

Proteins carry out several important functions within the cell, including structural support, catalysis of metabolic reactions, and regulation of complex cellular processes. Examples of structural proteins are actin and myosin in muscle, and keratin in hair, nails, and hooves. Regulatory proteins include hormones, growth factors, and receptors.

Polysaccharides

Polysaccharides function as food storage molecules and as structural molecules. The two most important polysaccharide food reserves in higher organisms are starch and glycogen, both of which are polymers of glucose sugar. Structural polysaccharides include cellulose and chitin: cellulose is the major constituent of cell walls in plants, and chitin is found in the exoskeletons of insects and crustaceans.

Lipids

Another important organic molecule, although it is not classified as a macromolecule, is the lipid. Lipids encompass a diverse group of compounds that are all soluble in nonpolar, organic solvents. Included in this class are fats, which are used primarily for energy storage; phospholipids, which are found in cell membranes; sphingolipids, which are especially prominent in the cell membranes of brain and nervous tissue; glycolipids, which are important in the myelin sheath of nervous tissue; steroids, which include male and female sex hormones, bile acids, adrenocortical hormones, and cholesterol; and fatty acids, which are components of energy storage molecules.

Endoplasmic reticulum

Endoplasmic reticulum (ER) is contiguous with the outer membrane of the nucleus. It is the site for folding proteins and assembling large molecules in an oxidizing environment. The ER consists of membranous channels lacking ribosomes (smooth ER) or containing ribosomes (rough ER). In the rough ER, ribosomes actively synthesize protein that accumulates in the lumen of the ER. These proteins include secretory proteins that make their way to the cell surface via a complex route, e.g., the rough ER, Golgi complex (see later), and secretory vesicles. Smooth ER is the site of synthesis of lipids and steroids and also for the inactivation and detoxification of drugs and other compounds harmful to cells.

Golgi complex

The Golgi complex (or Golgi apparatus) is a region of flattened vesicles closely related to the smooth ER both in proximity and function. It processes and packages secretory proteins and synthesizes complex polysaccharides. The Golgi also accepts vesicles that “bud off” the ER. These vesicles and their protein contents are processed further and then passed on, via vesicle budding of the Golgi complex, to other components of the cell. Therefore, the Golgi complex is a processing station for both receiving vesicles that fuse with it and also producing vesicles from it in a repackaged form, usually ready for export from the cell.

Lysosome and peroxisome

Two structurally similar organelles are the lysosome and peroxisome, each contained by a single membrane. Lysosomes are storage structures for hydrolases, i.e., enzymes that digest food and cell components that are no longer needed. Peroxisomes generate and degrade hydrogen peroxide. Animal peroxisomes also detoxify other harmful compounds, such as, ethanol, methanol, formate, and formaldehyde, and generate some unusual substances, such as D-amino acids.

Mitochondria

The mitochondrion (plural: mitochondria) is quite large, relative to other organelles, i.e., several micrometers (microns) in length and 1 micrometer in width, about the size of a bacterial cell. All mitochondria in the cells of an individual are maternally derived from those that were present in the egg at the time of fertilization. Therefore, unlike nuclear DNA in which the paternal contribution of genes is 50%, all mitochondrial DNA comes from the mother.

Depending on the organism and cell type, a cell may have only one mitochondrion or it may have several thousand. A typical human cell contains hundreds of mitochondria, each with 2–10 copies of mitochondrial DNA (mtDNA), resulting in thousands of copies of mtDNA per cell. Again in contrast with nuclear DNA inheritance, where each cell receives exactly half the genetic material at cell division, mitochondria are not always evenly partitioned into daughter cells – one cell may receive more (or fewer) copies of mitochondria. Therefore, the number of mitochondria and constituent mtDNAs can be heterogeneous between tissues and even within a given tissue; this is termed *heteroplasmy*.

Within the mitochondria, oxidation of nutrients (oxidative phosphorylation) takes place, providing energy to synthesize adenosine triphosphate (ATP). ATP conserves the energy from the oxidative reaction that would otherwise have been lost as heat and makes it available to the cell for work. Thus, the mitochondria have been called the powerhouses of the cell. They have a double membrane, an outer membrane plus an inner membrane, which are infolded into numerous projections called *cristae*, where oxidation of nutrients takes place.

Mitochondria also command special interest because they contain their own DNA (mtDNA) and ribosomes, although the ribosomes in mitochondria are more similar to those in prokaryotes in size and nucleotide sequence than to ribosomes elsewhere in the eukaryotic cell. mtDNA is usually circular, like a bacterial genome, with no histones attached. Human mtDNA contains 37 genes, including those that specify transfer RNAs, ribosomal RNA, and polypeptides important in ATP synthesis. The mitochondrion even encodes some of its own RNA and polypeptides, about 5% of those it needs. Mutations of mitochondrial genes, even when only a fraction contain mutant mtDNAs, can cause disease if they are located in tissues where mitochondrial function is important; for example, mitochondrial mutations have been implicated in several metabolic diseases, heart disease, and aging.

Ribosomes

In addition to the membranous organelles mentioned above, cells contain other important structures. The ribosome, made up of 50–80 different proteins and three or four different kinds of RNA molecules, is a tiny spherical body on which the synthesis of proteins takes place. They are found either free in the cytoplasm or attached to mitochondria, ER, or the outer surface of the nuclear membrane. Proteins needed for use in the cytosol are usually synthesized on single ribosomes.

Polypeptide chains are made on groups of ribosomes called polyribosomes, or polysomes. The polysome contains a variable number of ribosomes held together by a messenger RNA (mRNA) strand. This mRNA strand determines the sequence of amino acids in the synthesized protein. Signals residing in the mRNA also determine initiation, elongation, and termination of the polypeptide. Antibiotics, such as streptomycin, chloramphenicol, and puromycin, block protein synthesis at one of these three stages.

Centrioles

Tubules and filaments are other versatile cell components. Electron microscopy has shown that centrioles (or basal bodies), which are important in cell division, are found near the nucleus. The centriole contains nine microtubule triplets around its periphery. These bodies occur in pairs, called a diplosome or centrosome, which are perpendicular to each other and are attached to the outside of the nucleus. During the G₁/S cell cycle transition the centrioles self-duplicate and migrate to opposite ends of the cell, where they form spindle fibers (also made of microtubules). Spindle fibers help separate chromosomes to their respective daughter cells in cell division.

Many proteins interact with and regulate the centrosomes. Alterations in these centrosome-associated proteins can have pathological consequences. For example, mutations in the TP53 gene can lead to extra copies of the centrosome, predisposing the cell to misshapen spindle apparatus formation, aneuploidy, and tumor formation [3].

In the laboratory, colchicine inhibits cells from completing mitosis by binding to the monomer tubulin, thereby blocking its assembly into polymeric spindle fibers. Colchicine also indirectly disassembles already-formed spindle fibers. Without spindle fibers, chromosomes are unable to move away from the metaphase plate and complete cell division.

Cilia and flagella

Cilia and flagella, the external hair-like projections that function in cell motility, are also made of microtubules. Nine doublet tubules are arranged around the periphery and a tenth doublet forms the core. Like their close relatives, the microfilaments, microtubules are involved in cell movement, cytoplasmic streaming, cell cleavage, and membrane invaginations. Microtubule-initiated motion almost always requires ATP as an energy source.

1.1.3 Nucleus

The nucleus is the information headquarters for the cell. Unlike prokaryotes, such as bacteria and blue green algae that carry their genetic material in the cytoplasm, other more complex organisms confine their genetic information, DNA, within a nucleus. These more evolved plant and animal organisms, including humans, are termed eukaryotes (eu = true; karyon = nucleus). Every eukaryotic cell has a nucleus at some stage of its existence. Some cells have more than one nucleus, and some, such as red cells and platelets, lose their nuclei when they mature. Cells lacking nuclei, however, are severely limited in their metabolic activities.

The nucleus contains a nuclear membrane, chromatin, and nucleoli (see Figure 1.1). It is also the site of ribosome precursor assembly. The term *nuclear matrix* refers to the fibrous material that remains if the chromatin and nucleoli are extracted. DNA within the nucleus determines the cell's morphological, biochemical, and metabolic characteristics.

The appearance of the nucleus is markedly different in interphase (nondividing) and mitotic (dividing) cells. First noted by Brown in plant cells in 1831 [4], the interphase nucleus is a conspicuous spherical body in the cell interior. By light microscopy, it appears as an amorphous network of variably condensed fibers, called chromatin, which are not distinguishable as individual entities. Highly condensed chromatin stains darkly with nuclear stains and is known as *heterochromatin*; the more dispersed chromatin, which stains lightly or not at all, is called *euchromatin*. In cell division, the chromatin condenses into deeply staining, threadlike or rod-like structures called chromosomes (chromo = color; soma = body), which are present in specific numbers in each cell of a given species. This process of chromatin condensation to form chromosomes during division is necessary for the equal parceling of genetic information to daughter cells.

The nucleus is spatially organized with each chromosome in a specific region. This serves to prevent one chromosome from getting tangled with another. Telomeres are attached to the nuclear membrane. Between chromosomal subcompartments are chromatin-free interchromosomal domains. Here reside RNA molecules being processed for export to the cytoplasm. Highly transcribed portions of the chromosomes are positioned next to the interchromosomal domains and since different genes are transcribed in different cell types; the arrangement can vary from cell to cell.

Under the electron microscope, chromatin and chromosomes appear as fibrous structures. This is understandable since they comprise DNA molecules that are themselves filamentous. Fibers of DNA with associated proteins are about 30 nanometers (nm) in diameter, but protein-depleted strands are only about 10 nm in diameter. Chromatin fibers with diameters greater than 30 nm are occasionally observed and are believed to represent coiling or folding of these main fibers.

Nuclear envelope

The nuclear envelope, as the membrane surrounding the nucleus is called, is a porous double membrane with ribosomes attached to the outside. Numerous pores serve as channels for water-soluble molecules to travel between the nucleus and the cytoplasm. Ribosomes, mRNA, chromosomal proteins, and enzymes needed for nuclear activities are also thought to travel through these nuclear pores. The outer membrane is contiguous with the ER at many sites. Inside the nucleus are two obvious structural elements – the nucleolus and the chromatin. During cell division, this nuclear envelope disappears.

Nucleolus

One to four nucleoli appear as darkly staining bodies eccentrically placed within the normal nucleus. They comprise primarily RNA and protein but contain some DNA. Their size will vary, based on the cell type and the metabolic state, i.e., larger nucleoli are seen in rapidly dividing cells and in cells active in protein synthesis. Each nucleolus is formed along the nucleolar-organizing region (NOR) of one or more specific chromosomes; these regions are recognizable during cell division. The nucleolus is the site of ribosome precursor assembly; therefore, all ribosomes in the cytoplasm originate in the nucleolus.

Nucleic acids

The nucleus contains the nucleic acids DNA and RNA along with structural and regulatory proteins. Nucleic acids are involved with protein synthesis and the storage of genetic information. There are two kinds of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), each a polymer of *nucleotides*. Nucleotides consist of one purine or pyrimidine, a five-carbon sugar, and a phosphate group (see DNA). The sugar in DNA is deoxyribose; in RNA (see RNA) it is ribose.

DNA is the genetic material, and RNA is responsible for carrying out the instructions coded by the DNA. The primary functions of nucleic acids are *gene replication*, i.e., the process of copying sequences of DNA (genes) for distribution to daughter cells, and *gene transcription*, i.e., the process of copying sequences of DNA into complementary strands of RNA. These transcript RNAs may then be translated into corresponding sequences of amino acids during the synthesis of polypeptides (proteins). As previously discussed (see Ribosomes), protein synthesis occurs on cytoplasmic ribosomes.

DNA

The story of how scientists searched for the hereditary material and eventually established that DNA is the genetic material in almost all organisms is a fascinating one. Gregor Mendel's "hereditary factors," Walther Flemming's chromosomal threads, and Walter Sutton's chromosome theory of heredity led the way. Johan Miescher discovered DNA in 1869, calling it nuclein. The beauty of its structure and the logic of the coding process still inspire those who study them today.

Studies with sister chromatid exchange, electron microscopy, and other techniques demonstrate that a chromatid, one of a pair of metaphase chromosome strands, contains a single, uninterrupted, highly folded molecule of DNA. DNA itself is a double helix made up of two strands. Each strand is comprised of nucleotides, each consisting of a sugar molecule, a phosphate group, and one of four bases: adenine (A), guanine (G), thymine (T), or cytosine (C). The nucleotides are arranged side by side, with two bases forming one rung of a twisted ladder, and the phosphate and sugar form the outer structure (Figure 1.2). The sugar in

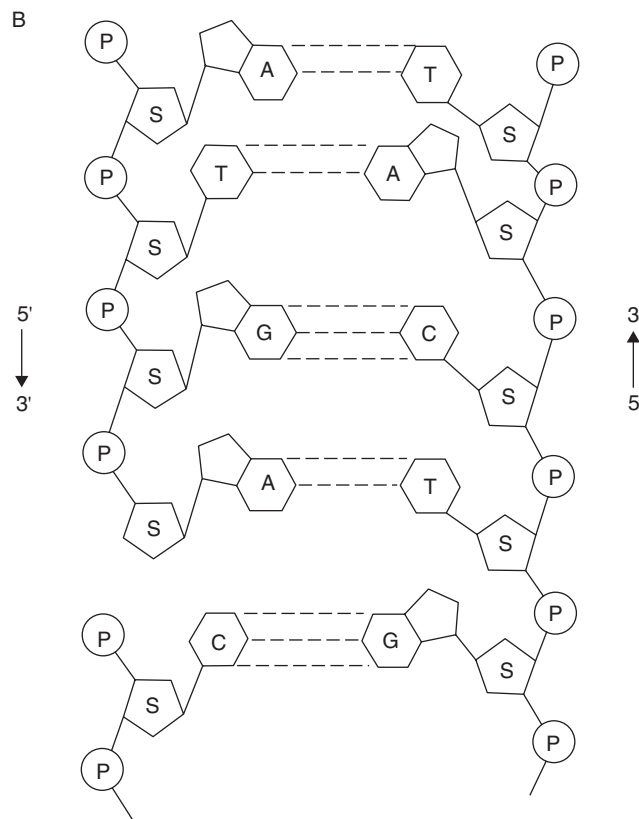


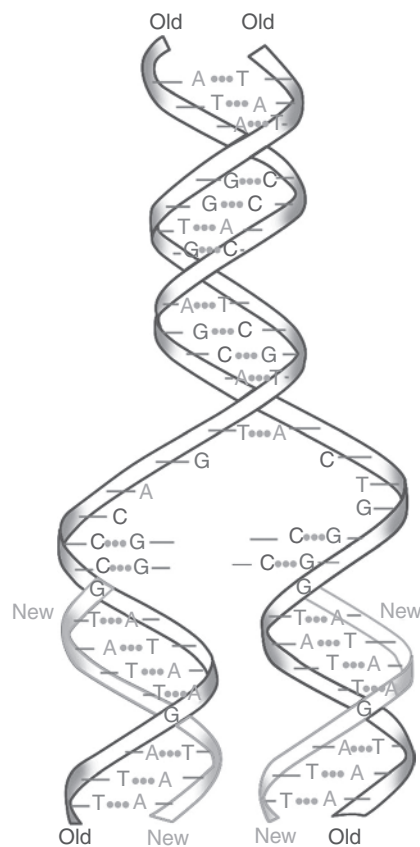
Figure 1.2 Chemical structure where A is adenine, T is thymidine, G is guanine, C is cytosine, S is sugar, and P is phosphate. The left strand polarity is from the 5' base to the 3' base, and the right strand has a 3' to 5' opposite polarity.

DNA, deoxyribose, has five carbon atoms, the third and fifth of which are bonded together by a phosphate (phosphodiester) linkage. Thus, a single strand of DNA is a polymer of deoxyribonucleotides held together by a 3'–5' phosphate linkage between their sugars. This is called the sugar–phosphate backbone of the DNA molecule, and it lies on the outside of the DNA fiber; the bases extend inward from the backbone. The free 3' and 5' ends give the molecule a polarity, or direction.

Watson and Crick [5,6] determined the double helical structure of DNA in the 1960s using models and X-ray diffraction images. The two strands of polynucleotides have opposite polarity. The bases hold the two strands together by hydrogen bonds (see Figure 1.2). Both strands are coiled in the same direction, so they cannot be separated without unwinding. Minor bases present in mammalian DNA include 6-methyl adenine and 5-methyl cytosine; the latter is found throughout the human genome but is often concentrated in areas of heterochromatin, such as in chromosomes 1, 9, 15, 16, and Y.

The bases in DNA are flat molecules that can stack on top of one another. The double-helical nature of DNA is maintained by these stacking forces and by the hydrogen bonds between the bases. The regularity of the double helix along its axis is possible because an AT pair is the same size and shape as a GC pair.

Prior to cell division, new DNA must be synthesized with great fidelity. This is accomplished by separation of the two strands so that each acts as a template for the assembly of a complementary strand (see Figure 1.3). Thus, two identical copies of the original DNA are produced, each composed of one original strand and one newly synthesized strand (semiconservative replication). This mechanism for producing a faithful copy of the genetic information for each daughter cell is fundamental to understanding techniques such as sister chromatid exchange (see Chapter 6, Chromosome stains). Of the four bases, two are purines (A and G), and two are pyrimidines (T and C). The precise replication of DNA is possible because the pairing of bases is specific: A pairs with T, and G pairs with C. Thus, the sequence of bases in one strand specifies the bases and their order in the complementary strand.



DNA replication illustrating two new helices being replicated semiconservatively.

Figure 1.3 This diagram illustrates two new helices being replicated semiconservatively (Adapted from Lince-Faria et al. 2009 [26]. See insert for color representation of this figure.).

Table 1.1 Genetic code

DNA triplet	RNA triplet	Amino acid
AAA	UUU	Phenylalanine
AAT	UUA	Leucine
TAA	AUU	Isoleucine
TAC	AUG	Methionine (start)
AGA	UCU	Serine
GGA	CCU	Proline
TGA	ACU	Threonine
CGA	GCU	Alanine
ATA	UAU	Tyrosine
ATT	UAA	(stop)
GTA	CAU	Histidine
GTT	CAA	Glutamine
TTA	AAU	Asparagine
TTT	AAA	Lysine
CTA	GAU	Aspartic acid
CTT	GAA	Glutamic acid
ACA	UGU	Cysteine
ACT	UGA	(stop)
ACC	UGG	Tryptophan
GCA	CGU	Arginine
CCA	GGU	Glycine

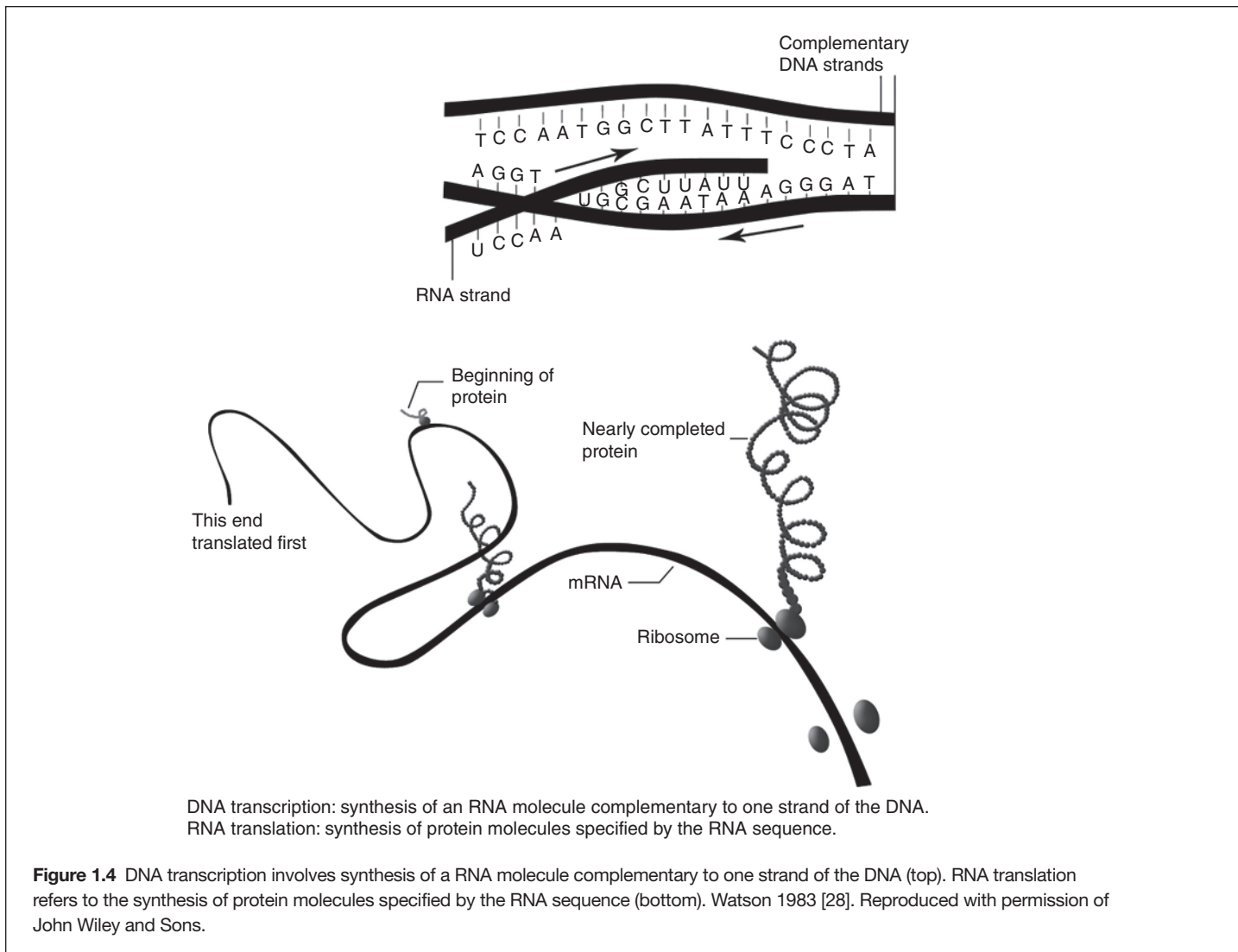
The nucleotide triplet in DNA specifies a triplet in RNA, which specifies an amino acid (or a start or stop signal). The code is "degenerate" in that each codon is not unique; for instance, UUA, UUG, CUU, CUC, CUA, and CUG all specify the amino acid leucine. A, adenine; C, cytosine; G, guanine; T, thymine; U, uracil.

The way DNA stores information was elucidated in the 1960s [5,6]. It was learned that the genetic code consists of three bases per code word; one triplet, or codon, codes for one amino acid (Table 1.1). A gene, then, can be understood as a linear arrangement of codons giving the instructions for the building of a protein with specific amino acids in a particular order.

It was later discovered that in higher eukaryotes, the coding instructions in a gene are often interrupted by DNA sequences that are not present in the mRNA and are not translated into amino acids in that gene's protein. These interrupting sequences are called introns (for intervening sequence), and the DNA sequences translated into the mRNA that usually code for protein are called exons [7]. The introns are spliced out of the mRNA before it codes for a protein. It is now known that a single gene can make more than one protein. Alternative splicing of introns can lead to multiple transcripts.

High temperatures or high pH conditions break the hydrogen bonds, and the double-stranded helix unwinds, or denatures, into two single-stranded helices. Because G–C pairs have three hydrogen bonds and A–T pairs have only two, the A–T pairs tend to be more unstable, denaturing before the G–C pairs. Therefore, the temperature at which a given DNA will be half denatured, or melted, is used as an index of the amount of G and C in that DNA. The curve of the rate at which this denatured DNA renatures (becomes double-helical once more) is called its Cot curve (Co = concentration of single stranded DNA, t = time). This curve yields other information about the DNA, such as how many sequences are present in multiple copies (repetitive DNA) versus how many are unique. Denaturation of DNA is an important step in fluorescence in situ hybridization (FISH) procedures discussed in a subsequent chapter.

Another measure of the G–C content is the buoyant density of the DNA. This is measured by forming gradients of concentration (and therefore of density) in cesium chloride during centrifugation. The DNA will collect at the



band where the gradient density is equal to the DNA density. This buoyant density depends upon DNA strandedness (single or double) and base composition. (See also 1.1.6, Satellite DNA.)

RNA

Like DNA, RNA is a polymer of ribonucleotides linked by 3'–5' phosphodiester bonds. RNA differs from DNA in three respects: its ribose sugar has a 2'-H group instead of 2'-OH group; it is single-stranded, rather than double-stranded; and it substitutes the base uracil for thymine to pair with adenine.

DNA does not specify a protein directly; rather, the gene for forming the protein is expressed through an intermediary molecule, the mRNA. Transcription, or mRNA synthesis, uses one strand of the DNA as a template for a complementary strand of RNA (Figure 1.4). After transcription, introns are spliced out and the mRNA molecule moves out of the nucleus to the cytoplasm, where it directs the synthesis of protein in the presence of ribosomes. Transfer RNA (tRNA) binds the appropriate amino acid to its anticodon, a base triplet complementary to a codon in mRNA. Ribosomal RNA assists in actual protein synthesis, by binding the anticodons of the tRNA molecules with the matching codons of the mRNA molecule so that the attached amino acids are covalently linked in the proper linear order.

Approximately 1.2% of the genome encodes for protein via mRNA; yet about 93% produces RNA transcripts. For example, micro RNA (miRNA) performs its regulatory function by binding to a matching region on a strand of mRNA; this will block

the ribosome from reading the strand and will thus disrupt protein synthesis. Likewise, small interfering RNA (siRNA) binds to mRNA and cuts it, thereby preventing translation. The functions of these and other RNAs are under study.

1.1.4 Chromosomes and their proteins

The two main categories of chromosome proteins are histones and nonhistones. Interphase chromatin contains mostly histone proteins, characterized by their basic pH, which is due to large numbers of the amino acids arginine and lysine. Their isoelectric points (pH at which the average charge of the molecule is zero) are always more than 10. Proteins with an isoelectric point less than 10 are classified as nonhistone proteins (NHPs). NHPs tend to be acidic, although their isoelectric points vary from 4 to 9, and are a mixture of proteins with different structural, enzymatic, and regulatory functions.

Histones

Histones, which can be found in a 1:1 ratio by weight to DNA, are classified into five major classes: lysine-rich H1, slightly lysine-rich H2A and H2B, and arginine-rich H3 and H4. More specialized forms can also be found in specific structures, such as, H5, the histone that replaces H1 in nucleated erythrocytes, and protamines, a group of highly basic proteins with a low molecular weight, which replace the histones in mature sperm. H3 and H4 have been highly conserved in evolution, and may actually express the same functionality in all eukaryotes [8]. Histones are also highly conserved in organisms from one tissue to another and between species; therefore, cows and peas have virtually the same histones. In the laboratory, histones can be extracted from chromatin by dilute acids or by high-molarity salt solutions. Acetic acid and methanol, commonly used to “fix” chromosomes, dissolve out some, if not most, of the histones.

Nucleosome

Histones H2A, H2B, H3, and H4 form octomers containing two molecules of each histone, giving rise to a 10-nm sphere or disk (visible with the electron microscope) called a nu-body, or nucleosome. The nucleosome appears to be the basic unit of eukaryotic chromatin. The fifth histone, H1, is implicated in the linking and compaction of these nucleosomes.

The nucleosome is present in dispersed or condensed chromatin, in repetitive areas or unique sequences, and in interphase and metaphase nuclei. It is associated with roughly 140 base pairs (bp) of DNA, which is wound twice around the spherical nucleosome. For reference, an average structural gene is approximately 1200 bp, which would span about six nucleosomes. When chromatin is extended by the removal of H1 histones, a linker region of about 60 base pairs of DNA can be seen between nucleosomes. When this region is uncoiled, the nucleosomes are seen located along the naked DNA like beads on a string. The H1 protein is responsible for condensing these beads into a 10-nm fiber. This is coiled again into the 25-nm strands, which look lumpy or kinky under the electron microscope (see Figure 1.5).

Nonhistone proteins

There are several hundred nonhistone proteins [9], which include all proteins of chromatin other than histones. Even though these proteins are thought to be more numerous and more variable than the histones, they actually make up much less of the chromatin mass [10]. Nonhistone proteins are involved in chromosomal metabolism [1,2], gene expression, and higher order structure.

Euchromatin and heterochromatin

Positive and negative heteropyknosis is used when referring to staining intensity that reflects the degree of coiling or condensation of the chromatin filaments. During the cell cycle, chromosomes condense and decondense, with maximum condensation at metaphase. Chromosomes and segments of chromosomes that are more heavily stained than the rest exhibit positive heteropyknosis. Others that are more lightly stained exhibit negative heteropyknosis. The chromatin in these variable regions, showing condensation unlike the remainder of the chromatin, is termed heterochromatin, while other regions are known as euchromatin. Under the electron microscope there is no difference in the basic structure of euchromatic and heterochromatic chromatin; therefore, differential staining has been attributed to fiber packaging within the heterochromatic regions. In the literature, the term heterochromatin usually refers to positive-heteropyknotic areas, but a given area may be negatively heteropyknotic in one banding technique and positive in another; therefore, use of the term should specify the stain with which it is being evaluated.

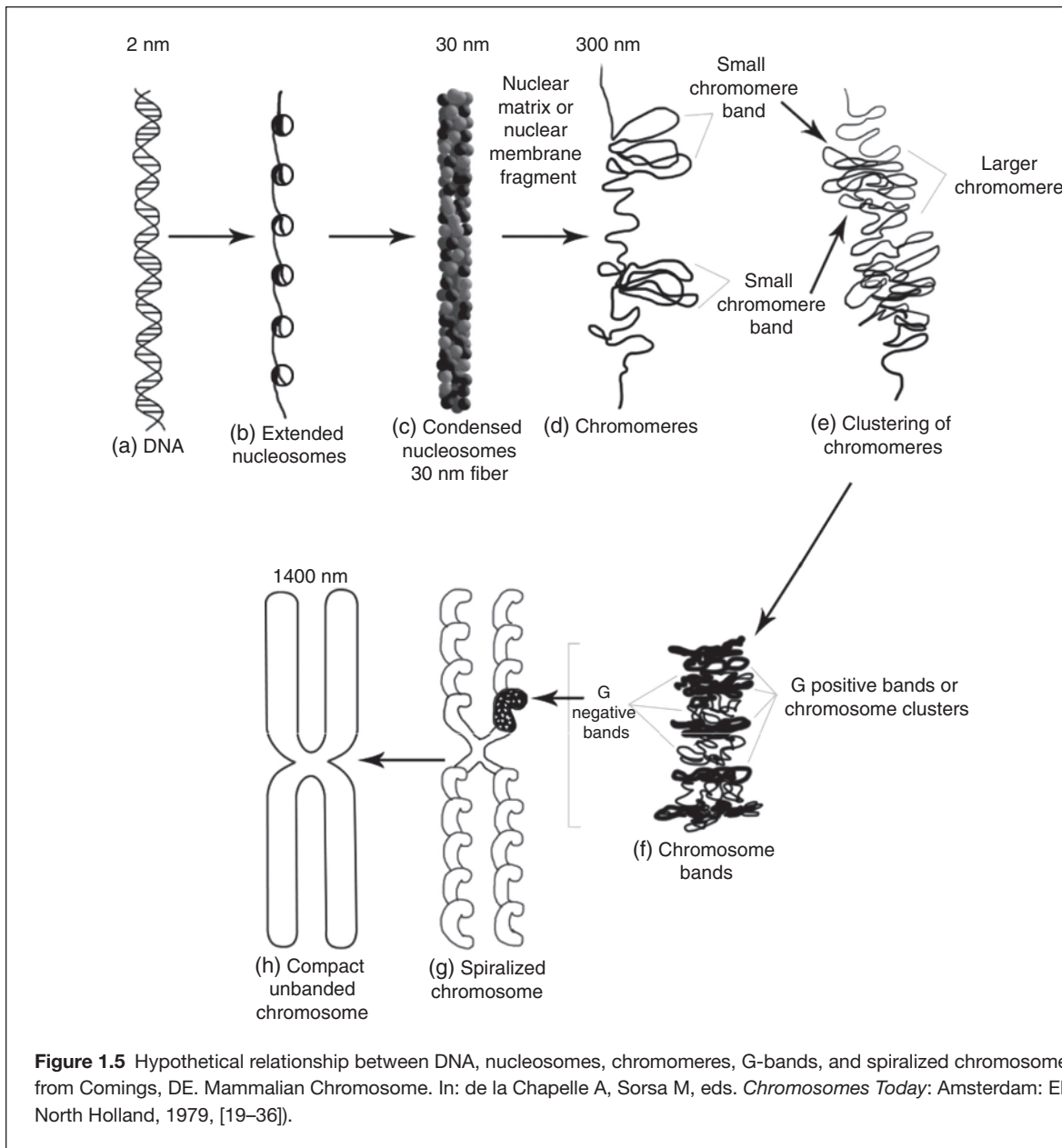


Figure 1.5 Hypothetical relationship between DNA, nucleosomes, chromomeres, G-bands, and spiralized chromosomes (Adapted from Comings, DE. Mammalian Chromosome. In: de la Chapelle A, Sorsa M, eds. *Chromosomes Today*: Amsterdam: Elsevier/ North Holland, 1979, [19–36]).

There are two distinct types of heterochromatin: facultative and constitutive. In humans, facultative heterochromatin is the name given to the condensed, inactive chromatin of X chromosomes in excess of one. It may represent one X in a given cell and its homolog in another, owing to the randomness of X inactivation (see 1.1.5, X inactivation). Constitutive heterochromatin is the name for the differentially staining areas of chromatin and chromosomes, which are evident with different stains and banding techniques and are constant from cell to cell.

Facultative heterochromatin and constitutive heterochromatin resemble each other in several ways: neither codes for protein (in most cases being genetically inactive), and both replicate late in the synthesis phase of the cell cycle. They differ in that constitutive heterochromatin is often rich in repetitive DNA, stains differently from euchromatin with banding techniques, and never elongates or decondenses. Facultative heterochromatin has sequences similar to active DNA, does not stain differently with standard banding techniques, and in some cases can become decondensed and active, as the X chromosome does during meiosis and early embryogenesis.

1.1.5 X inactivation

The manifestation of facultative heterochromatin in most female mammals, including humans, is a visible sex chromatin body that is derived from the second X chromosome. It is commonly referred to as a Barr body. Various staining techniques, as well as phase contrast, can be used to visualize the Barr body. It can be found in almost all tissue cells.

Barr body

Although earlier investigators noticed a “basophilic nucleolus” or “paranucleus,” this triangular staining body on the periphery of the nuclear membrane was not related to sex until 1949, when Murray L. Barr and Ewart G. Bertram published their observation of a “paranucleus” in cells from the hypoglossal nerve of female cats, but not of the males, called the Barr body [11]. Expanding on the work by Liane Russell, Mary Lyon in 1961 put forth what we now know as the *Russell–Lyon hypothesis* [12], which stated the following:

1. One of the two X chromosomes is inactivated in human females.
2. The inactivated X may have either maternal or paternal origin in a given cell of an individual, and the choice is random.
3. The inactivation occurs in early embryogenesis.
4. Inactivation is stable, and descendants of a cell with an inactive X inherit that same X in an inactive state except in the germ cells (see later).

X inactivation is often called dosage compensation, that is, a mechanism for producing equal amounts of gene products in females having two X chromosomes and males having only one. We know that this compensation is effected by some mechanism, because females homozygous for hemophilia A have clotting times similar to those of affected hemizygous males. Also, glucose-6-phosphate dehydrogenase (G6PD) levels are similar in normal females (most of whom are homozygous) and in normal hemizygous males [13].

This mechanism, however, is by no means universal. In *Drosophila*, females are the homogametic sex (XX) and males are the heterogametic sex (XY), just as in mammals. The difference is that in *Drosophila* both Xs are active (transcribed), but the single X in males produces comparable amounts of gene products. Birds, reptiles, and butterflies have no dosage compensation of Z-linked genes, and the homogametic sex (ZZ), which is male, produces twice the level of gene products as the heterogametic sex (ZW).

Because of the random inactivation of X chromosomes, female mammals are mosaics for the genes on the X chromosome [14]. This is demonstrated by coat color in tortoise-shell and calico cats, which are heterozygous for black and orange X-linked alleles. Another example is the expression of G6PD in women who are heterozygous for the trait. Clones from a single cell produce either the mutant or the wild-type enzyme, never both, but in a random sampling of cells, about 50% produce the wild type and 50% the mutant clones.

Inactivation appears irreversible, even in human/rodent hybrid cells in which the inactive human X is present alone against a rodent background. It also remains stable in cells maintained in culture for many generations [15,16]. Reactivation of the entire X chromosome takes place only in the oocytes at some time before meiosis, and both X chromosomes are transcribed. This has been supported by experiments in which the presence of the X-linked enzyme HPRT (hypoxanthine phosphoribosyl transferase) was found to be at levels twice as great in females as in males in the morula stage of mouse embryogenesis. X inactivation occurs in waves from day 3.5 to day 13 of development. Germ cell progenitor cells in females are inactivated by day 12 [15].

In mammalian cells, an inactivation center is believed to reside on the long arm of the X chromosome at band Xq13 (XIC). There is no known case in which this proximal Xq area, called the *c region*, is deleted in the X chromosome, leading to the presumption that two active X chromosomes would not be viable [15]. In metaphase cells, the inactive X often appears shorter than its active homologue, and is frequently bent in the proximal long arm [17]. In X chromosomes with two *c* regions, bipartite Barr bodies are seen in some cells.

DNA methylation is believed to play a role in maintaining this repression of the inactive X. The gene *XIST* (inactive X specific transcript) has been found to be active only on the inactive X chromosome [18,19]. This gene, located in the region of the X inactivation center (XIC), does not code for a protein but rather produces an RNA product that “coats” the X chromosome on which it is expressed (i.e., acts in *cis*), and recruits other factors, including those that result in histone deacetylation and methylation, with the overall result of transcriptional repression of genes on the same chromosome. Experimental treatment of cells with 5-azacytidine produces hypomethylation of DNA and partial reactivation of selected loci, but not of the entire X [15,16].

X inactivation and conversion to the heterochromatic state involves interaction between noncoding transcripts such as *Xist*, chromatin modifiers, and factors involved in nuclear organization. These produce changes in chromatin structure and in the spatial reorganization of the X chromosome [20].

In individuals having one normal and one abnormal X, the abnormal X is usually the inactive one. There are indications, however, that inactivation is initially random even in these individuals but that the cells with an active normal X survive [15,16]. In X;autosome translocations, the normal X is usually inactive, but minority cell lines have been demonstrated. In unbalanced X;autosome translocations (usually offspring of balanced translocation carriers), the translocation chromosome is also usually inactive, but what is interesting is that the inactivation may or may not extend into the autosome.

Some genes on the X chromosome escape inactivation [21]. It is known that DNA synthesis is not synchronous in the late-replicating X but that it starts around the centromere and is followed by the short arm and the proximal part of the long arm. Several of the genes escaping inactivation are found in the early-replicating regions of the otherwise late-replicating X and are thought to cause the abnormalities associated with extra X chromosomes [16].

In the laboratory, the inactive X can be identified by growing cells in bromodeoxyuridine (BrdU) for 40–44 hours, and adding thymidine 6–7 hours before fixation. After staining with Hoechst 33258, the late-labeling X will be bright when examined by fluorescent microscopy. Alternatively, cultures can be grown in the presence of thymidine and pulsed with BrdU, resulting in a pale-staining, late-labeling X.

1.1.6 Satellite DNA

Repetitive DNA found in constitutive heterochromatin is often called satellite DNA because much of it separates from main band DNA by density gradient centrifugation. Satellite DNA has come to mean any highly repeated sequence, whether separable by ultracentrifugation or not. The satellite bands originally described are called classical satellites I, II, III, and so forth (Table 1.2). A substantial portion of each fraction is made up of a single family of simple repeats designated by the Arabic numeral corresponding to the Roman numeral [14]. Other pure sequences may be designated by lower case Greek letters, which also relate to the fraction from which they were derived. Polymorphisms that have arisen from mutations can be detected by restriction endonuclease digestion and electrophoretic separation. *Consensus sequences* are ones that are substantially the same, differing by only a few bases.

Alpha (α) and beta (β) satellite DNA, as well as classical satellite DNA, are found at the centromeres of all human chromosomes. Alpha satellite DNA has a consensus sequence so that probes made from a mix of the α satellite probes can be visualized at the centromeres of all the human chromosomes [16] (Table 1.2). More specific α satellite probes can also identify the centromeric regions of specific chromosomes (see Chapter 16, Fluorescence in situ hybridization).

In contrast to the undispersed repetitive sequences found in heterochromatin, dispersed repetitive sequences are found throughout the genome. They are the short (<500 bp) interspersed elements (SINES) and long interspersed sequences (LINES) [16]. SINES contain cleavage sites recognized by the restriction endonuclease Alu1 and are located in the quinacrine pale bands. LINES have cleavage sites for L1 and are located in the quinacrine bright bands. Together the Alu1 and L1 families make up about a third of the total repetitive DNA [14].

Other satellites of interest are microsatellites, SSR (simple sequence repeats), and minisatellites, all of which are interspersed throughout the genome. Microsatellites are di- or trinucleotide tandem repeats and are highly polymorphic. SSR are 3- to 6-bp repeats found in coding and noncoding DNAs and are also highly polymorphic. Minisatellites have longer repeats, usually more than 10 bp, and are usually located at the distal ends of chromosomes. These satellite DNAs are useful for DNA fingerprinting because of their highly polymorphic nature and are also useful as in situ probes [22].

Table 1.2 Characteristics of satellite DNA

Type of satellite DNA	Location	Length	Repeat sequences
Alpha satellite DNA	All human chromosome centromeres; Yq	0.17/0.34 kb	Varies between chromosomes
Beta satellite DNA	Chromosome 1, 9, 13, 14, 15, 21, 22, Yq	50–300 kb	68-bp monomers
Satellite I Unit A	Most human chromosomes	0.04 kb	17 bp
Satellite I Unit B	Most human chromosomes	25–48 bp in arrays	(A-B-A-B-A) ⁿ
Satellite II	Most chromosomes	0.05 kb	(GGAAT) ⁿ
Satellite III	Chromosomes 9 and 15, also most others	0.05 kb	CAACCCGA ⁿ /G ₆ T(GGAAT) ⁿ

1.2 The cell cycle

1.2.1 Interphase

The transition from interphase to cell division (mitosis) and back to interphase is called the cell cycle; therefore, when the cell is not dividing, it is said to be in interphase. This is the time when the nucleus is metabolically active. The nucleus is spatially organized with each chromosome in a specific region. This separation serves to prevent one chromosome from getting tangled with another. Telomeres are attached to the nuclear membrane. Between chromosomal subcompartments are chromatin-free interchromosomal domains, where RNA molecules are being processed for export to the cytoplasm. Highly transcribed portions of the chromosomes are positioned next to the interchromosomal domains and since different genes are transcribed in different cell types, the arrangement can vary from cell to cell.

Originally, it was assumed that the cell synthesized new DNA just previous to cell division. With the use of autoradiography and Feulgen staining techniques in the early 1950s; however, DNA synthesis was found to take place hours before any sign of mitosis, during interphase [23]. Furthermore, it is now known that synthesis of new DNA does not take place all at once. During the synthesis (S) period of 6–8 hours, some DNA will replicate early, and other DNA will replicate late. A given part of the chromosome, however, will almost always replicate at a certain time in the S period; for example, the inactivated X will always replicate late, as well as most of the dark-staining G-bands, and the light-staining G-bands will replicate early. The late-replicating portions of the chromosomes are usually considered genetically less active than the early-replicating areas. The discrepancy in replication timing between regions (and even whole chromosomes in the case of the X chromosome homologues) provides the basis for the banding technique called replication banding (see Chapter 6, section 6.3.2, Replication banding), which is most useful in discriminating the active from the inactive X chromosome.

There are four distinct stages in the cell cycle. Each phase is driven by cyclin-dependent kinases (CDKs) in conjunction with regulatory subunits called the cyclins. Cyclins are negatively regulated by cyclin-dependent kinase inhibitors (CKIs) such as p16, p21, and p27. CDKs phosphorylate target proteins at different stages of the cell cycle. For example, the retinoblastoma protein, pRb, associates with a transcription factor called E2F when pRb is hypophosphorylated. This association physically prevents E2F from dimerizing with its partner protein DP and thus inhibits the cell's passage into the synthesis phase. When pRb is phosphorylated by CDKs, it no longer binds E2F, and the cell cycle is allowed to progress through the synthesis phase. Mutations in the *RBI* gene are responsible for the eye cancer called retinoblastoma. Another cell cycle regulatory protein, TP53, can likewise stall the cell cycle at the G1/S transition to allow time for DNA repair or to initiate programmed cell death (apoptosis) when the damage is severe. Indeed, cell cycle regulation is intimately tied to DNA damage response: it is crucial that the cell have a series of checkpoints to prevent damaged DNA from being replicated, but checkpoints also exist at other points in the cell cycle. Alterations in any of these elements (e.g., cyclins, CDKs, pRb, *ATM*, *TP53*, and many others) have been observed in many human cancers, and several have their own cancer syndrome associated with alterations (e.g., TP53 mutations and Li–Fraumeni syndrome, and *RBI* and retinoblastoma) [3].

The time intervals for the four stages are consistent within a given cell type but vary between cell types. The four stages of the cell cycle are G1 (gap one), S (synthesis of DNA), G2 (gap two), and M (mitosis). The histone proteins of the chromosomes are, like DNA, synthesized during S. When the cell is in a resting stage and is not cycling, as when nutrients are scarce, it is said to be in G0.

The average mammalian cell cycle spans 18 hours, with a range of 12–24 hours, and a typical schedule for the portions of the cycle is nine hours for G1, 5 hours for S, 3 hours for G2, and 1 hour for M (see Figure 1.6). However, early embryonic cells can complete the cycle in 30 minutes, and an adult liver cell may take 1 year. These differences are due to shorter or longer G1 and G2 stages. If the cell is arrested at some point in S or G2, it cannot undergo cell division. During G2, the final preparations for mitosis are completed, and unless protein synthesis is inhibited, the cell will divide.

Different drugs inhibit the cell cycle at various stages. S inhibitors include amethopterin (methotrexate), hydroxyurea, and cytosine arabinoside. Naturally occurring mitotic arrestants include Vinca alkaloids, colchicine (also an alkaloid), and podophyllin.

1.2.2 Cell division

It has been firmly established since the time of Louis Pasteur that cells arise from pre-existing cells [24,25]. In order to donate genes to progeny, the parent, whether it is a single cell or multicellular organism, duplicates its genetic material (e.g., chromosomes) and transfers the copies to the offspring. Diploid organisms possess chromosomes in pairs called homologues, one inherited maternally and one inherited paternally. These pairs of chromosomes are normally similar or identical in shape, with a given gene found in a specific position (locus) on each chromosome of a pair. Even though these homologous genes determine similar functions or characteristics, they often are not identical. There are commonly alternate forms (alleles) of a gene on each homologous chromosome that code for two different expressions of the same gene, such as body color in fruit

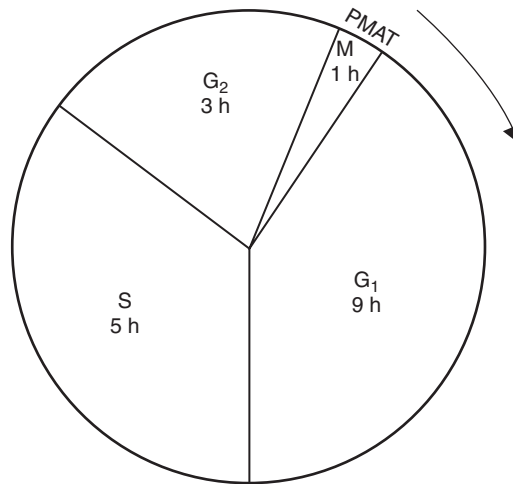


Figure 1.6 This pie graph demonstrates the four major stages of the human cell cycle and each stage's relative timeframe, e.g., Gap 1 (9 hours), Synthesis (5 hours), Gap 2 (3 hours), and Mitosis (Prophase, Metaphase, Anaphase, Telophase) (1 hour).

flies. If a fly inherits the allele for yellow from both parents, it is said to be homozygous for that trait. If, instead, it receives one allele for yellow and one for gray, it is said to be heterozygous for the trait. The color it expresses (gray or yellow) is called its phenotype, and the description of the actual genes it is carrying is its genotype.

In the case of the sex chromosomes in humans, males have only one X chromosome and the genes on that X are present unpaired, in a single dose. This condition is neither homozygous nor heterozygous, but hemizygous. Dosage compensation (see 1.1.5, X inactivation) provides a mechanism for the production of comparable gene products between homozygous and hemizygous individuals.

The genetic information present in each cell is transmitted during cell division, both in mitosis and meiosis. Mitosis is the division of somatic (nonreproductive) cells, and meiosis is the specialized division that occurs only in the formation of gametes (ova and sperm). One reason for the existence of two types of cell division is the fact that somatic cells need a full (diploid) complement of chromosomes; whereas, gametes need only one-half a complement (haploid), because they will ultimately fuse with another gamete during fertilization to become diploid in the resulting zygote.

1.2.3 Mitosis

In humans, the diploid ($2n$) number of chromosomes is 46. During the S stage of the cell cycle, each chromatid replicates itself. The cell at the end of S phase is still diploid, but the DNA content doubles as the chromatids are replicated to produce a chromosome comprised of two chromatids. Thus, at the end of the S phase, the cell is $2n$, but $4c$, where c = DNA content. At mitosis, the chromosomes line up at the metaphase plate and one chromatid from each chromosome goes to each daughter cell. The cycle continues and the chromatids are then replicated at S phase. Mitosis can be divided into four stages: prophase, metaphase, anaphase, and telophase (see Figure 1.7).

Prophase

Prophase is the stage of progressive coiling of the already doubled chromosomes, which appear long and threadlike. In middle to late prophase the chromosomes can be seen as discrete units, each containing two chromatids and a centromere. During prophase, the nucleolus becomes undetectable under the light microscope. With the electron microscope it is apparent that the nucleolus becomes dispersed throughout the nucleus. In the cytoplasm, the centrioles (one pair of which has budded off of the other) start to migrate to opposite poles of the cell and to form the microtubules that make up the mitotic spindle.

Between prophase and metaphase (in a stage often called prometaphase), the nuclear envelope breaks down in most organisms, releasing the chromosomes into the cytoplasm, which is contained by the cytoplasmic membrane. The chromosomes move rather erratically toward the equatorial plane (metaphase plate) of the cell. At this time the spindle fibers are not yet attached to the spot in the centromere (called the kinetochore) to which they later anchor for chromatid separation.

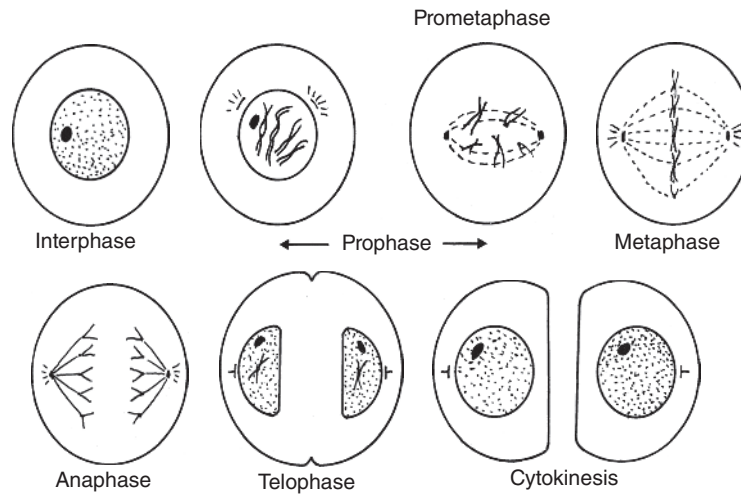


Figure 1.7 During the 1 hour that mitosis typically takes, replicated chromatin condenses in prophase to form identifiable chromosomes, and the nuclear envelope breaks down. Chromosomes that line up at the metaphase plate are attached at their kinetochores with spindle fibers that are connected at the other end to the centrioles at each of the poles. At anaphase, the spindle fibers pull the duplicated chromosome arms to opposite poles to form the two daughter cells in telophase, when the nuclear envelope is re-formed. Cytokinesis is the formation of new cell membranes between the daughter cells, each with a complete, identical set of genetic information.

Metaphase

At metaphase, the mitotic spindle is complete: the centrioles are in place at opposite poles, the chromosomes are lined up at the metaphase plate with the spindle fiber attached to each kinetochore, and the kinetochores are facing opposite poles of the cell. The mitotic spindle consists not only of these chromosome-to-pole fibers but also of continuous pole-to-pole fibers, which bypass the chromosomes altogether. At metaphase, chromosomes are at their most contracted state. This is also their least metabolically active state.

Anaphase

Anaphase begins with the division of the centromeres and, therefore, the separation of the chromatids. Once separate, each chromatid is called a daughter chromosome. The chromosomes move to opposite ends of the cell along the spindle fibers. Increased understanding of the mechanism of chromosome movement is being gained and some think there is potential to exploit this knowledge in the development of new cancer therapies [26,27].

Telophase

The final stage of mitosis is telophase. In telophase, the chromosomes uncoil, the nucleolus reappears, the nuclear envelope reappears, the spindle fibers disappear, and the nucleus takes on the morphology of the interphase cell. During or directly after telophase, the cytoplasm is divided by the formation of new cell membranes in a process called cytokinesis, and cell division is complete. The result of mitosis is two daughter cells, each with a complete, identical set of genetic information. For a video of mitosis in a live cell, search on “mitosis” or “cell division” on YouTube (www.youtube.com). A good example is found at: <http://www.youtube.com/watch?v=aDAw2Zg4IgE>.

1.2.4 Meiosis

Meiosis is often called the reduction division, because it reduces the number of chromosomes in each daughter cell to the haploid (n) number, which is 23 in humans. Meiosis takes place in the reproductive organs – the ovaries in females and the testes in males. The process of meiosis transforms cells called primary spermatocytes in the male testis and primary oocytes in the female ovary into haploid spermatids and ova, respectively. When fertilization occurs, the ovum and sperm fuse to form a

diploid zygote. Meiosis differs from mitosis in its reduction to four separate n (haploid) nuclei and in the creation of new gene combinations by crossing over so that the daughter chromosomes are composites of the parent chromosomes. The parent cell, for example, contains a pair of chromosomes number 1, one from each of its parents. During meiosis, these homologs exchange genetic material between them (crossing over – see later) so that the spermatid or ovum receives a single chromosome 1 that is derived from, but is not identical to, either of the two chromosomes 1 that the parent possessed. The same sequence of events has simultaneously occurred for chromosomes 2, 3, and so on. The consequence of this feature of meiosis is an increase in the phenotypic variation of sexually reproducing organisms, which provides selective advantage of great importance. The mechanism of this genetic exchange will become apparent in the description of meiosis.

Meiosis has two nuclear divisions, meiosis I and meiosis II. In meiosis I the homologous chromosomes (homologues) separate, and in meiosis II the chromatids separate (as in mitosis). This results in four cells, each with one haploid set of chromosomes. In the female, only one of these becomes a viable ovum, and the rest become polar bodies. In the male, all four spermatids can mature into spermatozoa.

Compared with mitosis, meiosis is a complicated and lengthy process. Like mitosis, each stage of meiosis has a prophase, metaphase, anaphase, and telophase. Prophase I is especially complex and is divided into five consecutive substages: leptotene (also called leptonema), zygotene (zygonema), pachytene (pachynema), diplotene (diplonema), and diakinesis (see Figure 1.8).

Prophase I

Leptotene

At this substage the nuclear chromatin begins to condense for division, but the chromosomes are not yet evident and the double nature of the strands cannot be discerned. The electron microscope shows that both ends (telomeres) of the chromosomes are attached to the nuclear envelope. By light microscopy, the leptotene cell has an enlarged nucleus and finely dispersed chromatin. Once the cell has entered leptonema, it is committed to meiosis.

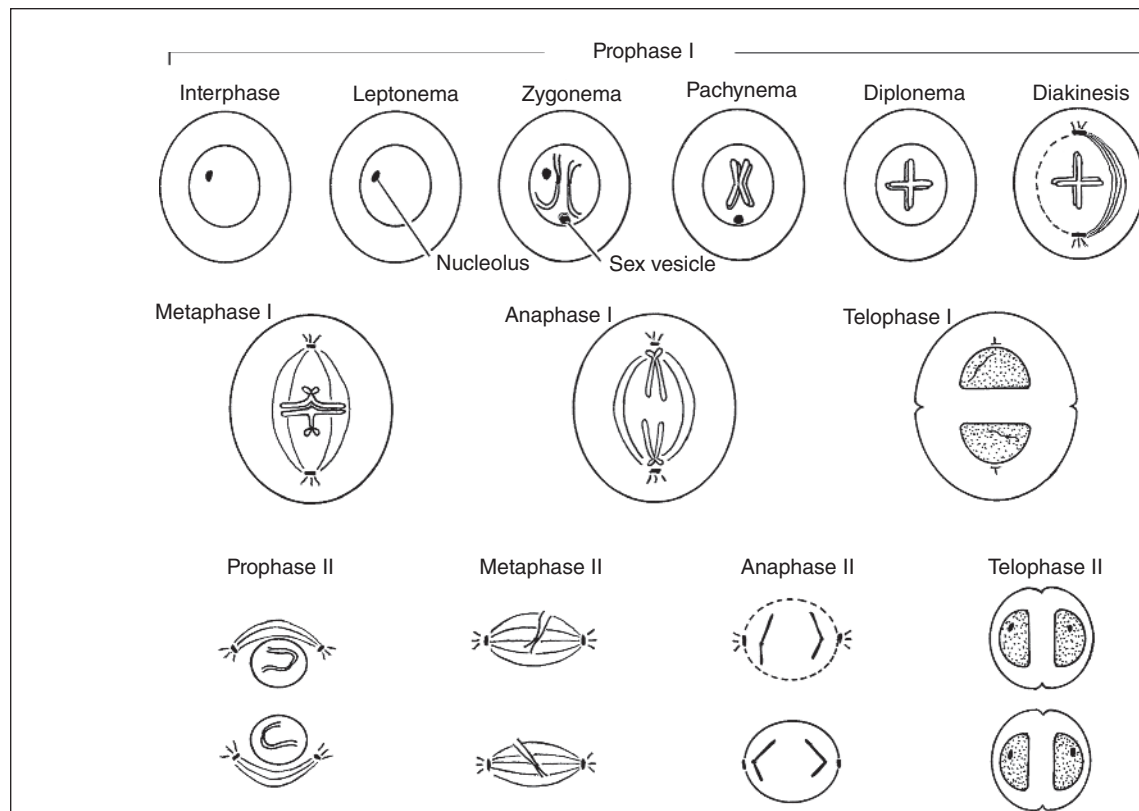


Figure 1.8 Meiosis, which takes place in the reproductive organs, starts with diploid cells and ends with haploid germ cells, i.e., spermatids and ova. Genetic crossing over, critical for genetic variation, occurs in prophase I. A second cell division, meiosis II, is responsible for the reduction of the diploid count to two haploid sets of chromosomes.

Zygotene

Homologous chromosomes, which still resemble long threads, align themselves side by side and attach to each other, allowing homologous loci to lie next to each other. This alignment is called synapsis (the Greek word for joining together). Electron microscopy reveals a structure holding the chromosomes close to each other; this structure is called the synaptonemal complex. The exchange of genetic material probably takes place in these complexes. This exchange is called crossing over, or recombination. In the female, the two X chromosomes are homologous and behave exactly like the other chromosomes. In males, however, the X and Y chromosomes are not homologous and condense to form a small, dark-staining body called the sex vesicle. Evidence shows that the X and Y are aligned end to end in this vesicle. Each set of homologues in synapse is called a bivalent, because the two chromosomes do not yet appear differentiated into chromatids. During zygotene the nucleolus is still visible and is associated closely with some of the bivalents. Zygonema ends when all homologues have been paired.

Pachytene

During this stage of prophase I, the bivalents shorten and become very thick, and crossing over occurs. Two nonsister chromatids cross over, but the other two remain unchanged. In good preparations, the bivalents can be seen as two parallel strands (chromosomes) on which there are a number of dark-staining, beadlike structures called chromomeres. These chromomeres are constant from preparation to preparation, so their number and position can identify some of the bivalents. Chromomeres are thought by some workers to be analogous to the G-band patterns of mitotic prometaphase chromosomes (see Figure 1.5). In this stage, the bivalent, owing to its four closely opposed chromatids, is known as a tetrad.

Diplotene and diakinesis

At the diplotene stage the nucleolus detaches from its associated bivalents, and the bivalent chromosomes begin to separate as their centromeres pull them apart. They are still attached, however, at points called chiasmata (a single point is a chiasma), which are the sites of genetic crossing over. Chiasma means cross, and chromosomes having only one chiasma typically assume a cross-like appearance. Each chiasma acts as a sort of knot that holds the paired chromosomes together so that the chromatids do not separate. In normal human meiosis, there is usually at least one chiasma per bivalent. If no chiasmata are present, nondisjunction can occur, leading to aneuploidy. In males, diplonema marks the disappearance of the sex vesicle, and the continued end to end association of X and Y can be observed.

Diakinesis occurs at this point in males: the chiasmata appear to move toward the ends of the bivalents (terminalize), the nucleolus dissipates, and the nuclear envelope disappears. In human females, however, meiosis is halted before diakinesis in the ovaries of the unborn female fetus. The oocytes remain in this special diplonema stage, called dictyotene, as long as 50 years or until each oocyte is singly ovulated after puberty before reaching the diakinesis stage. Meiosis is never completed unless fertilization occurs. When diakinesis is finished, the cell is ready to move into metaphase I.

Metaphase I

Here the spindle is formed and the bivalents line up at the equatorial plane. In females, the spindle is off-center in the cell and determines, by its position, which of the daughter cells will inherit most of the cytoplasm at anaphase. The bivalent chromosomes are lined up so that maternally and paternally derived chromosomes randomly face one pole or the other. This will allow them to sort independently to the daughter cells; each daughter cell will contain a mix of both paternal and maternal chromosomes. Excluding crossing over, this shuffling of 23 pairs of chromosomes can produce about 8 million genetically different gametes.

The form of the individual bivalents in metaphase I depends upon the number of chiasmata present. Bivalents with one chiasma form a cross-shaped structure at diakinesis, which proceeds to form a rod-shaped structure at metaphase when the chiasma terminalizes. Two terminalized chiasmata form a ring-shaped structure at metaphase; three chiasmata create a figure eight; and four or more will appear with additional loops. The X and Y chromosomes sometimes appear separated in metaphase I and are then called univalents.

Anaphase I and telophase I

Whole chromosomes, centromeres intact, move to opposite ends of the cell during these stages. In the oocyte of the female, one of the two daughter cells receives most of the cytoplasm and becomes the secondary oocyte. The other cell, mostly nucleus, becomes the first polar body. In humans, the cycle proceeds directly to meiosis II without an intervening interphase stage.

Metaphase II

In human meiosis there is no true prophase II; the cells pass directly into the second meiotic metaphase. The 23 chromosomes, each composed of two chromatids, move to the equatorial plate (metaphase II). At this stage, the chromosomes appear somewhat spiralized and fluffy. Although only a haploid set of chromosomes is present, i.e., one of each human homologue, there is still a diploid amount of DNA because the replicated strands have not yet separated. In anaphase and telophase, the two chromatids finally separate and go to two daughter cells so that the end product of meiosis is four haploid cells, each with one complete but different set of genetic material. In females, the spindle is again off-center, giving rise to a very large cell called an ovum and a second polar body. The first polar body may also undergo meiosis II, creating two additional polar bodies. Of the resulting four haploid cells, the ovum is theoretically the only viable gamete. For a video demonstration of the processes of meiosis, search 'meiosis' on YouTube (www.youtube.com). For example, this URL may be helpful to grasp the basic concepts: http://www.youtube.com/watch?v=R_LUJSqeSrl.

1.3 Recombinant DNA techniques

A breakthrough in the study of DNA occurred when researchers discovered bacterial enzymes that cut DNA at specific sequences [28]. These restriction enzymes often make staggered cuts in the two DNA strands, leaving short, single-stranded tails on the ends of both fragments. These single-stranded ends easily bind to complementary fragments by base pairing. Two fragments that have attached in this way can be permanently joined by adding DNA ligase, a repair enzyme that produces a recombinant molecule.

1.3.1 Bacterial-plasmid cloning

Further advances in this field took advantage of the fact that many bacteria contain plasmids, tiny circular DNA molecules, and that these plasmids can replicate autonomously in bacteria. Plasmids from the bacterium *Escherichia coli* (*E. coli*), which had only one recognition site for the restriction enzyme EcoRI, were cut by the enzyme; foreign DNA, also cut with EcoRI, was spliced in; and the plasmids were sealed with ligase. The hybrid plasmids were then transferred back into *E. coli*, where they carried out the instructions of the inserted DNA and reproduced with the bacteria's own DNA [29]. Using these techniques, researchers were able to isolate bacteria that had acquired a gene of interest and then make an enormous number of copies (cloning), owing to the rapid reproductive rate of bacteria [8]. This technique found practical uses; for instance, large amounts of insulin could be made for use by patients with diabetes.

Other advances in the study of DNA and genetics were made when Sanger, Maxam, and others devised methods for determining the base sequence of a given DNA molecule [30,31]. This capability led to the sequencing of mitochondrial DNA and bacteriophage λ . Sanger's dideoxy method was eventually used to sequence the entire human genome. Because restriction enzymes cut DNA at specific nucleotide sequences or recognition sites, the length of each fragment produced depends on the distance from one recognition site to the next. Harmless natural variations exist, such as the one at a point about 7000 nucleotides away from the beta-globin gene on chromosome 11. A recognition site for the restriction enzyme HpaI is present at that point in the DNA of some people, but not others. If the site is present, a short fragment containing the beta-globin gene, 7600 bp long, is produced. If absent, the beta-globin-negative fragment is 13,000 bp long. These normal variations have been named restriction fragment length polymorphisms (RFLPs).

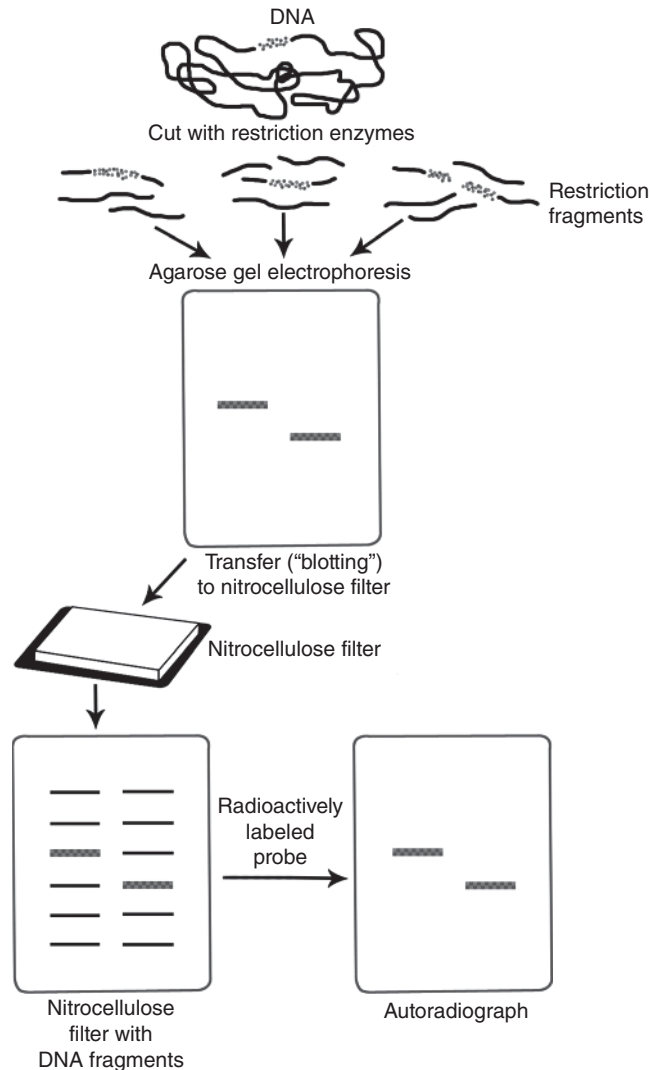
1.3.2 Electrophoresis

Fragments formed by restriction enzymes can be further separated by electrophoresis, a process in which the DNA fragments move through porous agarose gels (for large fragments) or polyacrylamide gels (for small fragments) under an electric field. Smaller fragments migrate more quickly, providing a visual way for determining fragment sizes by their positions in the gel. This separation comparison has proven useful in the study of heritable diseases like sickle cell disease [32].

Direct detection of the sickle cell globin gene has been demonstrated using the restriction enzyme MstII, which cuts within the beta-globin gene, as well as many other places. In the normal beta-globin gene, MstII cuts at the sequence CCTGAGG, producing two fragments of 1150 and 200 bp. The sickle cell mutation changes the sequence to CCTGTGG, thereby eliminating the recognition site; therefore, MstII produces only one 1350 bp fragment [28].

1.3.3 Southern blotting

Detecting whether a person has the two smaller fragments (normal) or the single 1350 bp fragment in sickle cell disease testing is complicated by the fact that there are so many fragments of similar sizes. This is overcome by using radioactively labeled probes in a technique called Southern blotting (see Figure 1.9).



Southern blotting. DNA fragments are separated according to size by agarose gel electrophoresis. They are then transferred to a nitrocellulose filter where they are exposed to radioactive probes that hybridize with complementary sequences. The radioactive signals are detected by autoradiography using x-ray film.

Figure 1.9 DNA fragments are separated according to size by agarose gel electrophoresis. They are then transferred to a nitrocellulose filter where they are exposed to radioactive probes which hybridize with complementary sequences. The radioactive signals are detected by autoradiography using X-ray film. Watson 1983 [28]. Reproduced with permission of John Wiley and Sons.

A probe is a short sequence of purified DNA that is complementary to the DNA of interest. In order for the probe to attach, the double-stranded DNA must be denatured by heat or alkali. The probe used to detect the sickle cell gene is a fragment of the cloned beta-globin gene, made radioactive so that it can be detected in the presence of large amounts of nontarget DNA.

DNA to be tested for the sickle cell mutation is cut with *Mst*II. The resulting fragments are then separated by agarose gel electrophoresis and treated chemically to denature them. Next, the fragments are transferred (blotted) onto a nitrocellulose filter, to which they become bound. The transfer retains the pattern of fragments that was produced in the agarose gel. Next, the filter is exposed to the radioactively labeled probe. Once hybridized to its complementary DNA, the radioactivity can be detected by placing the filter next to X-ray film, which exposes the film to produce an autoradiogram [28,32].

If a band corresponding to a DNA fragment of 1350 bp shows up on the autoradiogram, it represents the sickle cell gene. If two bands of 1150 and 200 bp appear, they represent the normal gene. If bands representing both the longer and the two shorter fragments are present, the individual has inherited the sickle cell gene from one parent and the normal beta-globin gene from the other parent and is a carrier for the sickle cell trait [28].

1.3.4 Synthetic oligonucleotides

Because it is unusual for a genetic mutation to exist in a restriction site, other means of detection must be used. One option is to use synthetic oligonucleotides (oligo=few). These short molecules are engineered to match portions of a normal gene exactly. If there is a change in just one base, the hybrid molecule will be unstable and will denature easily. These oligonucleotide probes can therefore be used to detect genetic defects that involve a point mutation or change in a single base [28,32].

1.3.5 Polymerase chain reaction

Another useful tool available to molecular geneticists is the polymerase chain reaction (PCR), which allows small amounts of DNA or RNA to be amplified, producing millions, even billions, of copies [33]. This makes it possible to make, from tiny samples of DNA, amounts great enough to be analyzed using restriction enzymes, oligonucleotides, or direct sequencing. The low cost of this method, in addition to its flexibility and precision for amplifying a short stretch of DNA, has resulted in PCR testing largely supplanting the bacterial plasmid cloning method for many clinical and research applications.

1.4 The human genome

1.4.1 Genomic DNA variations

Genomic DNA variation can provide useful methods by which to identify individuals. For instance, RFLP analysis is sometimes referred to as fingerprinting. However, these variations may also play a role in human phenotypic variation and disease. The single nucleotide is the smallest unit of variation. Single nucleotide polymorphisms (SNPs), by definition, are present in at least 1% of the population and may account for a large fraction of human genetic variation. Recently, surprising observations regarding human genetic variation on a larger scale were made. First, alterations in DNA copy number of large segments of DNA were reported in normal individuals. These segments (copy number variations, or CNVs) ranged from kilobases to megabases of DNA. The size of these alterations was below the resolution of karyotypic analysis and above the resolution of most molecular techniques, thus their discovery relied on the technique called array comparative genomic hybridization (aCGH), which is a sensitive method to detect genomic DNA imbalances at a lower limit of statistical resolution of several kilobases. Fifty-six percent of the clones overlapped with known coding regions and some included one or more genes [34].

Further research has demonstrated that individuals carry their own unique CNV signatures, and that CNV databases from populations or ethnic groups carry variations and similarities. Similarity between groups with unusual phenotypic characteristics or disease, has led to some CNVs being recategorized as pathogenic rather than benign population polymorphisms. The DNA architecture lying immediately adjacent to a region is thought to convey susceptibility to rearrangements observed as CNVs. Certain DNA regions contain low copy repeats (LCRs) that may be in direct head-to-tail orientation or in inverted tail-to-tail orientation; these repeats may facilitate nonallelic homologous recombination (NAHR) during meiosis [35] resulting in duplications or deletions. Many of the chromosomal deletion/duplication syndromes share this etiology. For example, velocardiofacial–DiGeorge syndrome (VCF/DGS) is usually caused by a deletion that results from unequal crossover events between two 240 kb LCRs termed LCR22-2 and LCR22-4 on chromosome 22 long arm in band q11.2. These LCRs may also cause duplications in the same region. Similarly, LCRs flank the Charcot–Marie–Tooth/HNPP (hereditary neuropathy with liability to pressure palsies) region on chromosome 17p12. Duplications lead to Charcot–Marie–Tooth syndrome (MIM: 118220) and deletions cause HNPP (MIM: 162500). Depending on the size, CNVs may be observed using a number of different methods, including chromosome analysis, FISH, and chromosome microarray analysis. Although many larger CNVs have been described using chromosome and FISH analysis, most of the recent additions to the CNV databases, and much of what we know of CNV formation, have been uncovered using array technologies [36].

The sequencing of the human genome, first published in 2001, has opened up a new era in understanding genetics from a functional viewpoint [37,38]. One early surprise in this process was the relatively low number of genes resulting in a peptide or protein. The number is believed to be 21–30,000 rather than the expected 100,000. Only about 1% of the genome is comprised of exons. Introns account for another 25%.

A consortium of international scientists is seeking to identify all biologically functional elements in the genome (ENCODE project). The results of their study of the first 1% have revealed more surprises [39,40]. The majority of DNA is transcribed into RNA, and the transcripts overlap one another extensively, but about half are not constrained evolutionarily. A number of previously unknown start sites for DNA transcription have been identified. Regulatory sequences are symmetrically distributed around transcription start sites with no bias toward upstream regions. Chromatin accessibility and histone modification patterns are highly predictive of both the presence and activity of transcription start sites.

Although alternative splicing of transcripts has been known for some time, the ENCODE project has verified that splicing occurs for virtually all genes. Also, a given gene can code for more than one protein. The estimate is that the average protein-coding region produces 5.7 different transcripts. Exons may be spliced into transcripts from other genes, even from different chromosomes. So, while the absolute number of protein coding genes identified by the Human Genome Project was surprisingly low, through alternative splicing and post-translational modifications, the actual protein repertoire of the human genome is quite large. When this is superimposed on the different gene expression and protein interaction networks in particular cells and tissues, the 'protein function' repertoire becomes larger still.

References

1. Becker WM, Kleinsmith LJ, Hardin J. *The World of the Cell*, 6th ed. San Francisco, CA: Pearson Benjamin Cummings; 2006.
2. Lewin B, Lingappa VR, Plopper G. *Cells*. Sudbury, MA: Jones and Bartlett; 2007.
3. Pelengaris S, Khan M. DNA replication and the cell cycle. In: Pelengaris S, Khan M, Blasco M, eds. *The Molecular Biology of Cancer*, 2nd ed. Hoboken, NJ: Wiley-Blackwell; 2006, 109.
4. Brown R. On the organs and mode of fecundation in Orchideae and Asclepieadeae. *Trans Linn Soc Lond (Bot)* 1833; 16: 685–745.
5. Watson JED, Crick FHC. A structure for deoxyribose nucleic acids. *Nature* 1953; 171: 737–738.
6. Watson JD, Crick FHC. The structure of DNA. *Cold Spring Harbor Symp Quant Biol* 1953; 18: 123–131.
7. DuPrav EJ. *DNA and Chromosomes*. New York: Holt, Rinehart and Winston; 1970.
8. Pines M. *The New Human Genetics – How Gene Splicing Helps Researchers Fight Inherited Disease*. Bethesda, MD: National Institute of General Medical Sciences; 1984.
9. Elgin SCER, Amero SA, Eissenberg JC, Fleischmann G, Gilmour DS, James TC. Distribution patterns of non-histone chromosomal proteins on polytene chromosomes: functional correlations. In: Gustafson JP, Appels R, eds. *Chromosome Structure and Function, Impact of New Concepts*. New York: Plenum Press; 1988, 145–156.
10. Lewin B. *Genes IV*. New York: Oxford University Press; 1990.
11. Barr MEL, Bertram EG. A morphological distinction between neurons of the male and female, and the behavior of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature (London)* 1949; 163: 676–677.
12. Lyon MF. Sex chromatin and gene action in the X chromosome of mammals. In: Moore KL, ed. *The Sex Chromatin*. Philadelphia, PA: WB Saunders; 1966, 7–15.
13. Levitan M. *Textbook of Human Genetics*, 3rd ed. New York: Oxford University Press; 1988.
14. Wagner RP, Maguire MP, Stallings RL. *Chromosomes: A Synthesis*. New York: Wiley Liss; 1993.
15. Lock LF, Martin GER. Dosage compensation in mammals: X chromosome inactivation. In: Risley MS, ed. *Chromosome Structure and Function*. New York: Van Nostrand Reinhold; 1987.
16. Thurman E, Susman M. *Human Chromosomes: Structure, Behavior, Effects*, 3rd ed. New York: Springer-Verlag; 1993.
17. Flejter WL, VanDyke DL, Weiss L. Bends in human mitotic-metaphase chromosomes including a bend marking the X-inactivation center. *Am J Hum Genet* 1984; 36(1): 218–226.
18. Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 1991; 349(6304): 38–44.
19. Brown CJ, Lafreniere RG, Powers VE, Sebastio G, Ballabio A, Pettigrew AL, Ledbetter DK, Levy E, Craig IW, Willard HF. Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* 1991; 349(6304): 82–84.
20. Chow J, Heard E. X inactivation and the complexities of silencing a sex chromosome. *Curr Opin Cell Biol* 2009; 21: 359–366.

21. Brown CJ, Flenniken AM, Williams BRG, Willard HF. X chromosome inactivation of the human TIMP gene. *Nucleic Acids Res* 1990; 18(44): 4191–4195.
22. Verma RS, Babu A. *Human Chromosomes: Principles and Techniques*. New York: McGraw-Hill; 1995.
23. Mazia D. Mitosis and the physiology of cell division. In: Brachet J, Mirsky A, eds. *The Cell*, vol 3. New York: Academic Press; 1961, 77–412.
24. Conant JB. *Science and Common Sense*. New Haven, CT: Yale University Press; 1951.
25. Dubos RJ, Pasteur L. *Free Lance of Science*. Boston: Little, Brown and Company; 1950.
26. Lince-Faria M, Maffini S, Orr B, Ding Y, Florindo C, Sunkel CE, Tavares A, Johansen J, Johansen KM, Maiato H. Spatiotemporal control of mitosis by the conserved spindle matrix protein Megator. *J Cell Biol* 2009; 184(5): 647–657.
27. Wood KW, Chua P, Sutton D, Jackson JR. Centromere-associated protein E: a motor that puts the brakes on the mitotic checkpoint. *Clin Cancer Res* 2008; 14(23): 7588–7592.
28. Watson JD, Tooze J, Kurtz DT. *Recombinant DNA—A Short Course*. Scientific American Books. New York: WH Freeman; 1983.
29. Beaudet AL, Scriver CR, Sly WS. Genetics and biochemistry of variant human phenotypes. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. *The Metabolic Basis of Inherited Disease*, 6th ed. New York: McGraw-Hill; 1989, 3–53.
30. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA* 1977; 74: 560–564.
31. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 1975; 94: 441–448.
32. Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM. *Molecular Biology of the Gene*, 4th ed. Menlo Park, CA: Benjamin-Cummings; 1986.
33. White TJ, Arnheim N, Erlich HA. The polymerase chain reaction. In: Stewart A, ed. *Trends in Genetics*, Cambridge: Elsevier Trends Journals; 1989; 5(6), E185–188.
34. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. *Nature Genetics* 2004; 3(69): 949–951.
35. Lupski JR, Stankiewicz P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genetics* 2005; 1(6): e49.
36. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin NE, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TEC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. *Science* 2004; 305(5683): 525–528.
37. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos DE, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. *Nature* 2006; 444(7118): 444–454.
38. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferreira S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I,

- Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome. *Science*. 2001; 291: 1304–1351.
39. International Human Genome Sequencing Consortium. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D. Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860–921.
40. Birney D, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007; 447: 799–816.