1

Introduction to the Cell

In this chapter we will introduce the main properties of the eukaryotic cell, starting from its characterization in terms of its organelles, such as the nucleus, and its structural components, such as the cytoskeleton and the membrane (see Section 1.1). In Section 1.2, we discuss in detail how DNA is organized, and we introduce different chromatin structures such as B-DNA and Z-DNA and their condensation into chromosomes. Section 1.3 discusses how DNA is replicated so that the genetic information it encodes can be passed over to daughter cells. We also explain how DNA is repaired when damage due to external perturbation occurs. Next, in Section 1.4, we explain how the genetic information encoded in the DNA is transcribed into RNA and then translated into proteins, a process that has been termed the "central dogma" of molecular biology. Cells are surrounded by the plasma membrane formed by lipid bilayers, which also enclose the organelles and have a key role in intracellular and extracellular transport. This issue is illustrated in Section 1.5. Section 1.6 discusses the regulation of gene expression in the cell, and in particular that performed by miRNAs, a set of small RNA molecules. Finally, Section 1.7 illustrates the process of cell division and Section 1.8 discusses cell death and cell senescence.

1.1 Architecture of the Eukaryotic Cell

A cell is a small organized machine where DNA stores information, RNAs translate the message in protein language and proteins are the effectors. The ingredients needed to control the behavior of a cell are a mixture of biochemical and physical factors. Before discussing in detail how the machine functions, we first describe its general architecture. Cells in the human body can differ widely in terms of shape, size and function, but some general features are common to all cell types and are illustrated in Figure 1.1.

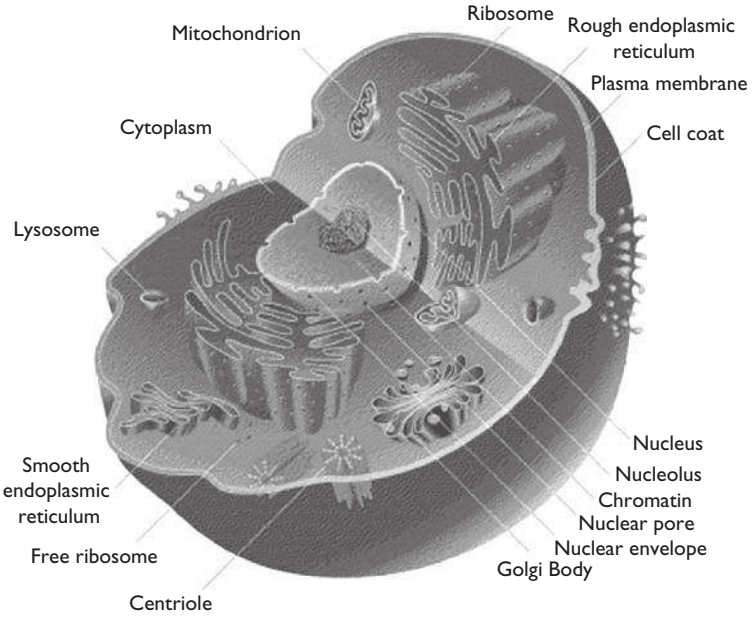

Figure 1.1 A schematic representation of an eukaryotic cell, displaying the organelles. Image by Mediran, CC-SA-BY-3.0 licence.

Eukaryotic cells are enclosed by the plasma membrane, a semi-permeable membrane in the form of a lipid bilayer that we will discuss in detail in Section 1.5. In contrast to bacterial cells, eukaryotic cells contain a set of membrane bound structures, known as organelles, that perform specialized functions. The largest cell organelle is the nucleus, which has a diameter that ranges between 3 to 10 μm and stores most of the cell DNA (Figure 1.2a). The nucleus is bound by a nuclear envelope composed of a protein network, enclosed in a bilayer membrane. Nuclear pores allow the passage of materials into and out of the nucleus into the cytoplasm which includes other organelles and the fluid cytosol.

The nucleus is surrounded by the endoplasmic reticulum (Figure 1.2c), a complicated maze of membrane-bound cisternae where ribosomes synthesize proteins. Proteins are shipped by vesicles to the Golgi apparatus, which looks like a set of layers of flattened disks and is the site of protein sorting (Figure 1.2d). From the Golgi apparatus, proteins are then transported to other regions of the cells. Mitochondria are shaped like cylinders with rounded ends, with a diameter of around 0.5 μm, and bound by a double membrane (Figure 1.2b). They are the site of production for adenosyne triphosphate (ATP), which provides the energy source for many cellular processes.

Cells display a complex mechanical structure provided by the cytoskeleton, a network of filaments of varying size and elastic properties (Figure 1.3); for

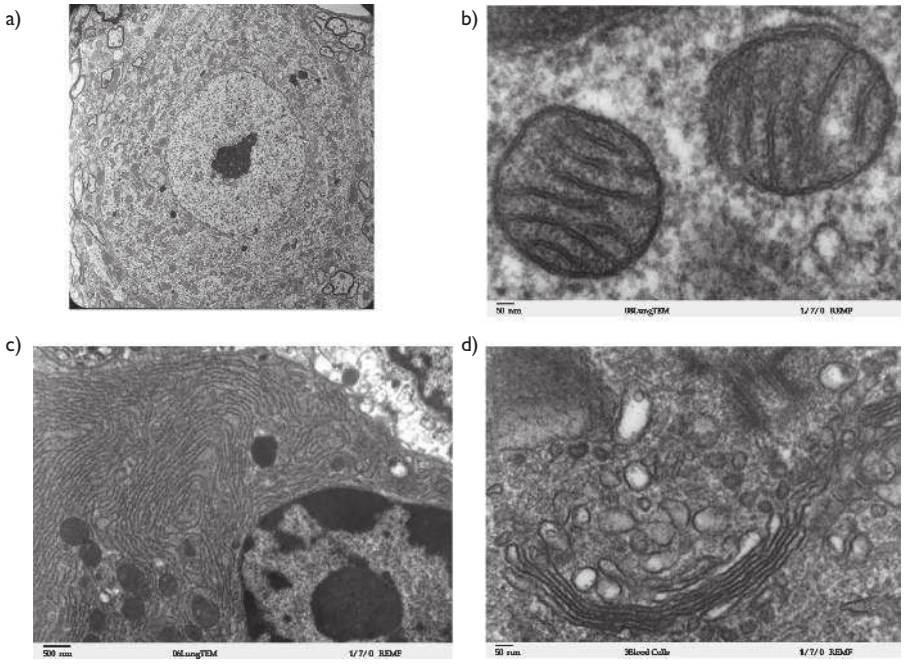

Figure 1.2 Cell organelles from transmission electron microscopy: a) A transmission electron micrograph of an eukaryotic cell (motor neuron from the bone marrow) showing the nucleus. Image by M. E. Pasini and M. Gioria. b) Mammalian lung cell showing mitochondria. c) Mammalian lung cell showing the nucleus and the rough endoplasmic reticulum. d) The Golgi apparatus from a red blood cell. Images b)–d) by Louisa Howard and Miguel Marin-Padilla, public domain.

instance, thin actin filaments, with diameter 7nm, often attached to the plasma membrane; a wide variety of intermediate filaments, with diameters around 10nm; and thicker and rigid microtubules, of diameter 25nm, radiating from the microtubule organizing center (MTOC). Microtubules play a key role in cell division.

1.2 The Organization of Genetic Material (DNA, Chromosomes, Genomes)

A DNA molecule is composed of two strands, held together by the hydrogen bonding between four bases: adenine (A), cytosine (C), guanine (G) and thymine (T). Adenine can form two hydrogen bonds with thymine; cytosine can form three hydrogen bonds with guanine. Although other base pairs (e.g., (G:T) and (C:T)) may also form hydrogen bonds, their strengths are not as great as those of (C:G) and (A:T) found in natural DNA molecules. Due to the specific base pairing, DNA's two strands are complementary to each other. Hence, the nucleotide sequence of one strand determines the sequence of the other strand. By convention, the sequence in

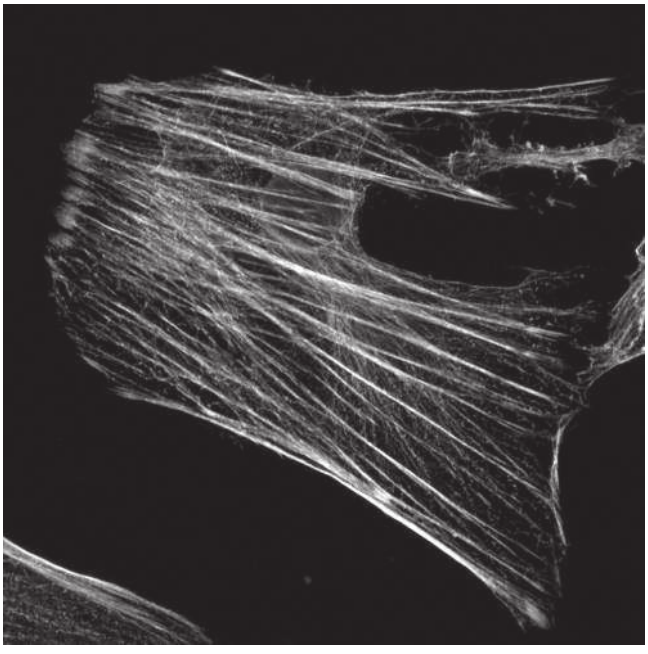

Figure 1.3 The cytoskeleton of HeLa cervical cancer cells. The fluorescent image shows actin filaments, microtubules and the cell nucleus. Image by S. Wilkinson and A. Marcus, National Cancer Institute, public domain.

a DNA database refers to the sequence of the 5′ to 3′ strand (left to right). Figure 1.4a shows the interaction between the nucleotides.

In a DNA molecule the two strands are not parallel, but intertwine with each other. Each strand looks like a helix and the two strands form a "double helix" structure, as discovered by James D. Watson and Francis Crick (Watson and Crick, 1953). In this right-handed structure, also known as the B form, the helix makes a turn every 3.4 nm, and the distance between two neighboring base pairs is 0.34 nm. Hence, there are about 10 pairs per turn. The intertwined strands form two grooves of different widths, referred to as the major groove and the minor groove, which may facilitate binding with specific proteins (see Figure 1.4b). In a solution with higher salt concentration or with alcohol added, the DNA structure may change to the A form, which is still right-handed, but making a turn every 2.3 nm with 11 base pairs per turn. Another DNA structure is called the Z form, because its bases resemble a zigzag. Z-DNA is left-handed (see Figure 1.4b). One turn spans 4.6 nm, comprising 12 base pairs. DNA molecules with alternating G-C sequences in alcohol or high salt solutions tend to have such a structure. The biological relevance of Z-DNA has attracted much attention in recent years. In fact, the Z-conformation was believed to trap the negative supercoiling when transcription occurs (Ha et al.,

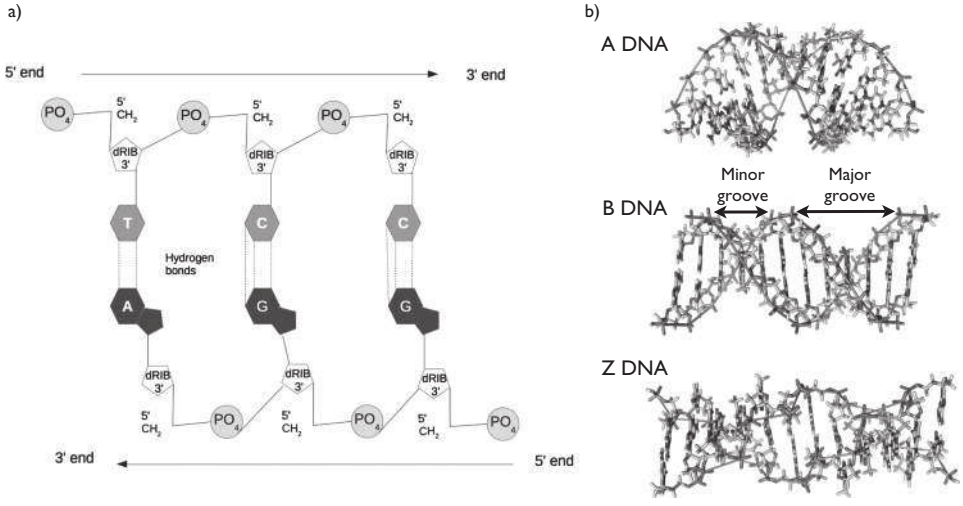

Figure 1.4 Structure of DNA. a) Interaction between nucleotides along DNA. b) DNA double helix in B and Z forms, differing by their chirality. The B form of DNA is characterized by an alternation between minor and major grooves. Adapted from an image by Richard Wheeler CC-BY-SA-3.0 licence.

2005; Rich and Zhang, 2003). Recent evidence shows a new role in gene transcription of the Z-DNA binding motif editing enzyme ADAR12529 (Kim et al., 1997). Moreover, Z-DNA is shown to be highly hydrophobic (Li et al., 2014).

The DNA duplex has two negatively charged phosphate backbone strands spiraling around the middle core of nucleotide base pairs (Figure 1.4b). Due to the high negative charge of this polyelectrolyte, the inter-DNA interaction is electrostatically repulsive. However, meters-long genomic DNA is packed by nature into compact structures in all living beings. To condense DNA, attractive forces must overcome repulsive forces. In most eukaryotic cell nuclei, DNA is packed in the form of chromatin by forming a complex with specialized histone proteins. Both analytical theories and computer simulations have been carried out to reveal the origin of like-charged DNA–DNA attraction and the roles of cations near DNA (Rouzina and Bloomfield, 1998; Shklovskii, 1999; Korolev et al., 2010). Generally, DNA–DNA interaction is caused by ion fluctuations and charge-bridging effects and is cation-dependent.

According to the most accepted model of DNA packing, chromosomes are hierarchically coiled: first, DNA is coiled around nucleosomes; second, the nucleosome strand is coiled into a 30 nm fibre; third, the 30 nm fibre is coiled into higher-order loops filling the volume of the chromosome (see Figure 1.5). A nucleosome consists of approximately 147 base pairs of DNA wrapped around eight histone protein cores. Linker DNA, more than 80 base pairs long, connects two histones between each nucleosome unit.

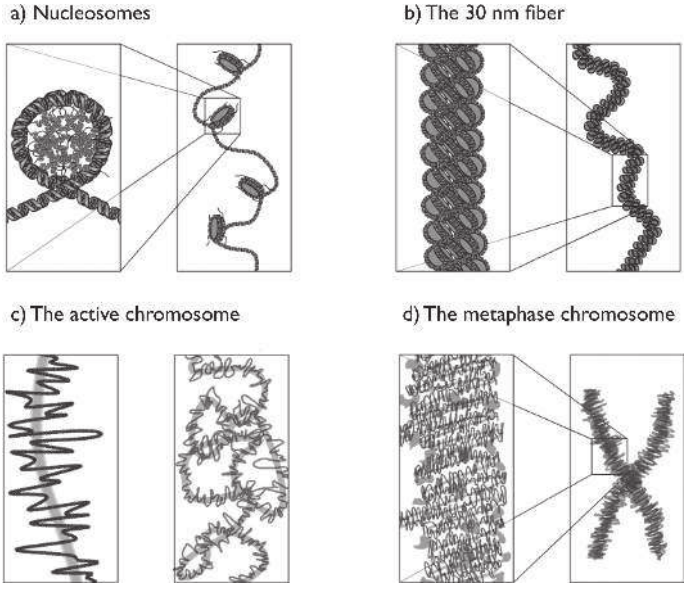

Figure 1.5 Structure of chromosomes. a) DNA is packed around histones, forming nucleosomes that are arranged like beads on a string. b) The nucleosomes are arranged in a 30 nm fiber. c) During interphase, chromosomes in the chromatin form long polymer-like structures thanks to scaffolding proteins. d) During cell division, in metaphase, chromatin condenses into chromosomes. Adapted from an image by Richard Wheeler, CC-BY-SA-3.0 licence.

The structure of the 30 nm fibre has not been resolved. There are two competing models proposed on the basis of in vitro data: the solenoid model and the zigzag arrangement of nucleosomes. In the one-start solenoid model, consecutive nucleosomes interact with each other and follow a helical trajectory with bending of linker DNA. In the two-start zigzag structure, two rows of nucleosomes form a two-start helix so that alternate nucleosomes become interacting partners, with relatively straight linker DNA. Twisting or coiling of the two stacks can produce different forms of the zigzag model. Recently, the existence of the 30 nm fibre has been questioned, due to the lack of reproducible observations in mammalian cells (Nishino et al., 2012), leading to alternative models (Fussner et al., 2011).

Artificial segments of chromosomes were created and tagged by the insertion of multiple copies of the lac operator sequence (Strukov and Belmont, 2009). The idea was to distinguish whether the same nucleosomes are always positioned in the same 3D space within a metaphase condensed chromosome. The results showed no reproducibility in the lateral positions of the tagged sequences in mitotic chromosomes, even in sister chromatids. These experimental results do not agree with a reproducible hierarchical folding model for the chromosome, suggesting instead a disordered compaction of chromatin at metaphase, or at some other stage in the

condensation process. In summary, the structure of chromatin is heterogeneous both locally and globally.

A proteinaceous chromosome scaffold was believed to anchor DNA loops. However, in 2002, Poirier and Marko measured the physical properties of chromosomes by holding them between two micropipettes, showing that chromosomes are elastic and that DNA contributes to the elasticity as much as the proteins (Poirier and Marko, 2002). These results preclude the possibility that chromosomes are shaped by a proteinaceous scaffold, and suggest instead that they are formed by a DNA–protein meshwork. These data were then confirmed by electron microscopy tomography of imaged snap-frozen chromosomes (König et al., 2007). Therefore, chromatin fibres criss-cross the chromosome and are interlinked at frequent intervals. The condensin protein complex is found at these interlinks, suggesting that this fundamental chromosome constituent may function as a crosslinker of the meshwork. More recently, small angle X-ray scattering showed that nucleosomes do not display a folding hierarchy (Nishino et al., 2012). The current view is that the chromosome is the result of a self-organizing nucleosome chain constrained by condensin-mediated interactions.

1.3 DNA Replication, Repair and Recombination

The information enclosed in the DNA should be passed to the daughter cells avoiding missing or changed content. The machine that duplicates DNA is called replication. From the biochemical point of view, DNA replication is an intricate process requiring the concerted action of many different proteins. Because each resulting DNA double helix retains one strand of the original DNA, DNA replication is said to be semi-conservative. We will not enter here into the details of the biological factors involved, since there are many books available on the subject, but focus instead on the open problems that lie at the boundary between physics and biology.

To begin the process of DNA replication, the two double helix strands are unwound and separated from each other by the helicase enzyme. The point where the DNA is separated into two single strands is known as the replication fork. This is also the place where new DNA will be synthesized. After the single strands are exposed, they are rapidly coated by single-strand binding proteins (SSBs). SSBs keep the strands separated during DNA replication, preventing them from snapping back together. SSBs are weakly bound to the DNA and they are displaced when the polymerase enzymes begin synthesizing new DNA strands. Once separated, the two single DNA strands act as templates for the production of two new, complementary DNA strands. Remember that the double helix consists of two antiparallel DNA strands with complementary 5′ to 3′ strands running in opposite

directions. Polymerase enzymes can synthesize nucleic acid strands only in the direction running from 5′ to 3′, hooking the 5′ phosphate group of an incoming nucleotide onto the 3′ hydroxyl group at the end of the growing nucleic acid chain.

When the strands are separated, DNA polymerase cannot simply begin copying the DNA. In fact, DNA polymerase can only extend a nucleic acid chain but cannot start one from scratch. To give DNA polymerase a place to start, an RNA polymerase called primase first copies a short stretch of the DNA strand. This creates a complementary RNA segment, up to 60 nucleotides long, called a primer. Now DNA polymerase can copy the DNA strand. The DNA polymerase starts at the 3′ end of the RNA primer, and, using the original DNA strand as a guide, begins to synthesize a new complementary DNA strand. Two polymerase enzymes are required, one for each parental DNA strand. Due to the antiparallel nature of the DNA strands, however, the polymerase enzymes on the two strands start to move in opposite directions. One polymerase can remain on its DNA template and copy the DNA in one continuous strand. However, the other polymerase can only copy a short stretch of DNA before it runs into the primer of the previously sequenced fragment. It is therefore forced to repeatedly release the DNA strand and slide further upstream to begin extension from another RNA primer. The sliding clamp helps hold this DNA polymerase onto the DNA as the DNA moves through the replication machinery. The continuously synthesized strand is known as the leading strand, while the strand that is synthesized in short pieces is known as the lagging strand. The short stretches of DNA that make up the lagging strand are known as Okazaki fragments.

The information contained inside DNA can be altered by mutations that could be induced by UV light, the most important source of DNA damage, or by the action of other chemical agents or errors during replication. These errors can be corrected by pre-replication or post-replication repair. Genetic information can be stored stably in DNA sequences only because a large set of DNA repair enzymes continuously scan the DNA and replace any damaged nucleotides. DNA repair mechanisms rely on the presence of a separate copy of the genetic information in each of the two strands of the DNA double helix. An accidental lesion in one strand can therefore be removed by a repair enzyme. Afterwards, a correct strand is resynthesized using the undamaged strand as a reference. Most of the DNA damage is removed thanks to one of two major DNA repair pathways. In base excision repair, the altered base is removed by a DNA glycosylase enzyme, followed by excision of the resulting sugar phosphate. In nucleotide excision repair, a small section of the DNA strand surrounding the damage is removed from the DNA double helix as an oligonucleotide. In both cases, the gap left in the DNA helix is filled in by the sequential action of DNA polymerase and DNA ligase, using the

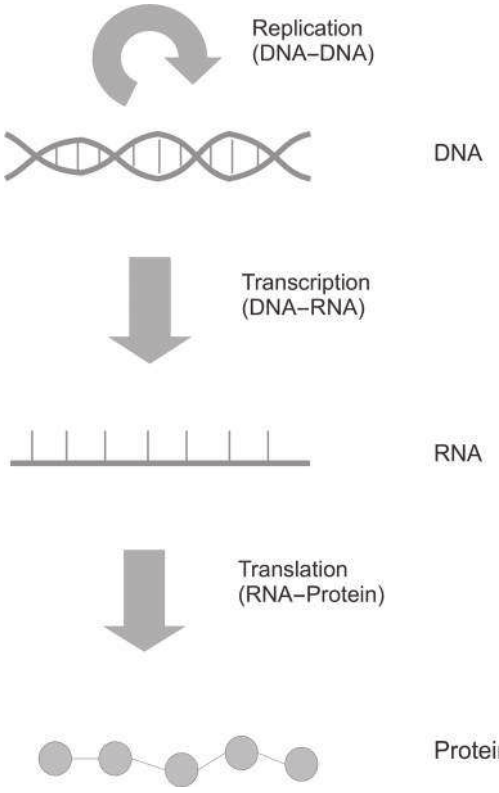

Figure 1.6 Illustration of the central dogma of molecular biology. DNA is both replicated into DNA and transcribed into RNA. RNA is translated into proteins.

undamaged DNA strand as the template. Other critical repair systems, based on either non-homologous or homologous end-joining mechanisms, reseal the accidental double-strand breaks occurring in the DNA helix. In most cells, an elevated level of DNA damage causes both an increased synthesis of repair enzymes and a delay in the cell cycle. Both factors help to ensure that DNA damage is repaired before a cell divides. In general the process is quite efficient and the rare errors result in specific non-common genetic pathologies such as the ataxiatelangiectasia (AT), whose symptoms include neurodegeneration, predisposition to cancer and genome instability.

Exchange of genetic information between two molecules of DNA leads to genetic recombination whose result is the production of a new combination of alleles. In eukaryotes, during meiosis, recombination contributes to exchange of DNA between people. Recombination can, however, also occur during mitosis involving the two sister chromosomes formed after chromosomal replication. Genetic recombination can also occur in bacteria and archaea.

1.4 Transcription and Translation Machineries

The information stored into the DNA should be translated into a message comprehensible for the cell, leading eventually to a specific phenotype. This role is played by the transcription and translation machineries. Transcription is the process by which the information in DNA is copied into messenger RNA (mRNA) for protein production. Finally, the translation machinery translates the mRNA language into proteins. It is important to remark that the amino acid code is degenerate and each triplet of nucleotides corresponds to more than a single amino acid. From this observation it follows that changes in the DNA sequence do not necessarily alter amino acid composition.

A detailed description of the transcription and translation machineries is easy to find in any biological textbook. Here we would like just to sketch a few interesting points. Transcription begins with a bundle of factors assembling at the promoter sequence on the DNA, until the arrival of the RNA polymerase that initiates transcription. These assembled proteins require contact with activator proteins that bind to specific sequences of DNA, known as enhancer regions. Once a contact is made, the RNA polymerase transcribes the gene by moving along the DNA. According to a well-accepted scenario, stable enhancer-promoter chromatin loops between regulatory regions (Tolhuis et al., 2002) would establish physical contacts with a promoter in static chromatin configurations, so that regulatory inputs are the same in all cells. Thus transcriptional control would result from the action of binding molecules (Tolhuis et al., 2002). Alternatively, enhancer-promoter contacts could be considered as random events occurring in a fluctuating structural environment (Fussner et al., 2011; Nora et al., 2012). This view implies that transcriptional regulation would be intrinsically heterogeneous, contributing to the observed cell to cell variations (Amano et al., 2009; Krijger and de Laat, 2013). Following this idea, a recent paper suggests that the contacts between potential regulatory elements occur in the context of fluctuating structures rather than stable loops (Giorgetti et al., 2014).

1.5 Membrane Structure and Intracellular Trafficking

The plasma membrane surrounding the cell is formed by a lipid bilayer, whose structure is revealed by high-magnification electron micrographs such as the one shown in Fig. 1.7a, showing two dense lines separated by a space (Cooper, 2000). The lines are due to binding to the polar head groups of the phospholipids of the metallic atoms used in the preparation of the sample for transmission electron microscopy. The electron-dense lines are separated by the electron-poor internal portion of the membrane, containing hydrophobic fatty acid chains.