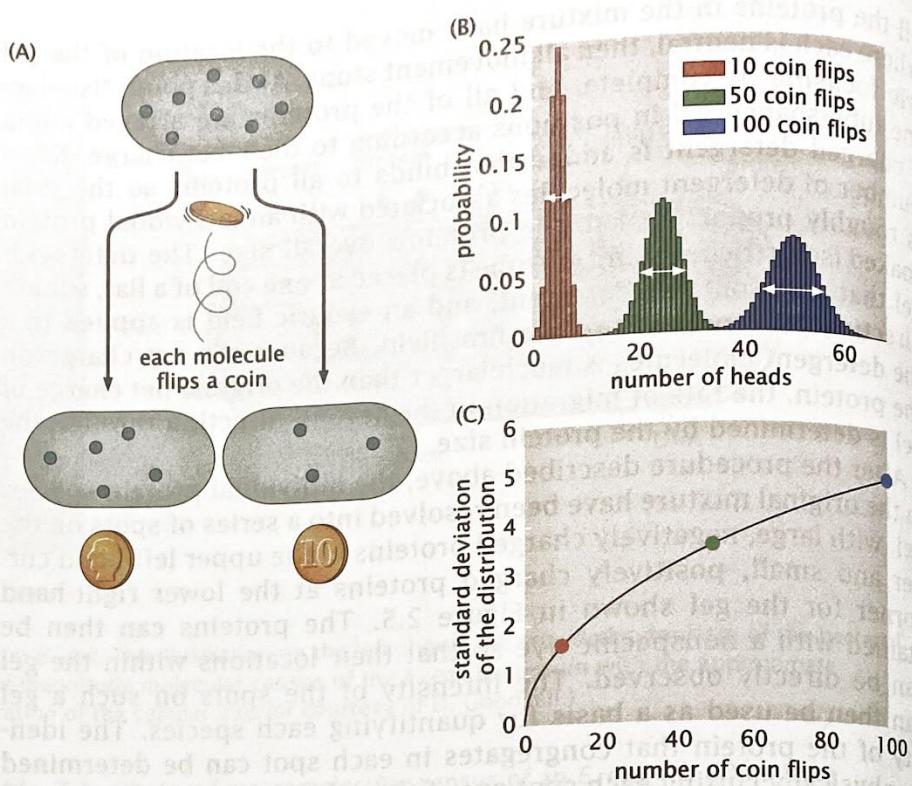


Figure 2.7: Schematic of the random partitioning of molecules during the process of cell division. (A) When the cell divides, each of the molecules chooses a daughter cell via a coin flip. (B) Probability distribution for the number of heads for different choices of the total number of coin flips. (C) The width of the distribution as a function of the number of coin flips.



of protein fusion in each strain of the library by comparing the total fluorescence in the cell with the single-molecule standard. Such a measurement for *E. coli* results in the histogram showed in Figure 2.6(B). It is important to point out, however, that these two methods are not in agreement, as shown in Figure 2.6(C). It is seen that, for a given protein, fluorescence microscopy tends to undercount proteins with respect to mass spectrometry (or mass spectrometry tends to overcount with respect to fluorescence microscopy). There are various possible sources for this discrepancy, ranging from systematic errors in the two experimental techniques to the fact that the experiments were done under different growth conditions leading to very different cell cycle times and cell sizes. One of the key features revealed by fluorescence measurements is the importance of cell-to-cell variability, which we can estimate using simple ideas from probability theory.

Estimate: Cell-To-Cell Variability in the Cellular Census

When cells divide, how much variability should we expect in the partitioning of the mRNA and protein contents of the different daughter cells? Conceptually, we can think of the passive molecular partitioning process as a series of coin flips in which each molecule of interest flips a coin to decide which of the two daughters it will go to. This idea is illustrated schematically in Figure 2.7.

To make a simple estimate of the variability, we turn to one of the most important probability distributions in all of science, namely, the binomial distribution. Despite the apparently contrived nature of coin flips, they are precisely the statistical experiment we need when thinking about cellular concentrations of molecules that are passively partitioned during cell division. The idea of the binomial distribution is that we carry out N trials of our coin flip process with the outcomes being heads and tails. However, a more biologically relevant



古法秦七圖

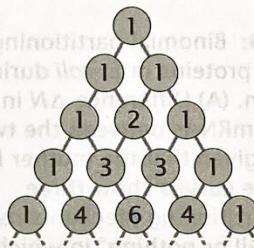
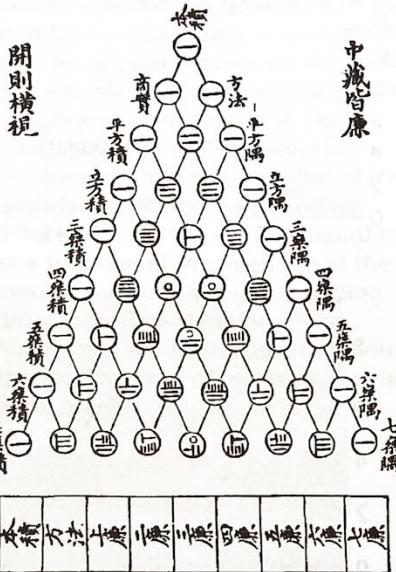


Figure 2.8: Binomial coefficients. (A) Though commonly known as Pascal's triangle, Chinese mathematicians had also earlier appreciated the symmetry of this collection of numbers. (B) The binomial coefficients that arise when raising $p+q$ to the power n , where n characterizes which row we are talking about, with the top "row" denoted by $n=0$.

way to think of these outcomes is that with each cell division, the macromolecule of interest partitions either to daughter 1 or to daughter 2, with the probability of going to daughter 1 given by p and the probability of going to daughter 2 given by $q = 1 - p$. If we are dealing with a fair coin, or if the sizes of the two daughters are equal and there are no active segregation mechanisms in play, then both outcomes are equally likely and have probability $p = q = 1/2$.

To be precise, the probability of having n_1 of our N molecules partition to daughter cell 1 is given by

$$p(n_1, N) = \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1}. \quad (2.6)$$

The factor $N!/[n_1!(N-n_1)!]$ reflects the fact that there are many different ways of flipping n_1 heads (h) and $N-n_1$ tails (t) and is given by the famed binomial coefficients (see Figure 2.8). For example, if $N=3$ and $n_1=2$, then our three trials could turn out in three ways as *thh*, *hth*, and *hht*, exactly what we would have found by constructing $3!/(2!1!)$. Note also that throughout this book we will see that, in certain limits, the binomial distribution can be approximated by the Gaussian distribution (large- N case) or the Poisson distribution ($p \ll 1$ case).

If the number of copies of our mRNA or protein of interest is 10, then on division each daughter will get 5 of these molecules on average. More formally, this can be stated as

$$\langle n_1 \rangle = Np. \quad (2.7)$$

For the case of a fair coin (i.e., $p=1/2$), this reduces to precisely the result we expect. This intuitive result can be appreciated quantitatively in the distributions shown in Figure 2.7(B). Here, the average number of heads obtained for a number of coin flips is given by the center of the corresponding distribution. The width of these distributions then gives a sense for how often there will be a result that deviates from the mean. As a result, to assess the cell-to-cell variability, we need to find a way to measure the width of the distribution. One way to characterize this width is by the standard deviation, which is computed from the function given in Equation 2.6 as follows:

$$\langle n_1^2 \rangle - \langle n_1 \rangle^2 = Npq, \quad (2.8)$$

a result the reader is invited to work out in the problems at the end of the chapter.

With these results in hand, we can now turn to quantifying the fluctuations that arise during partitioning. There is no one way to characterize the fluctuations, but a very convenient one is to construct the ratio

$$\frac{\sqrt{\langle n_1^2 \rangle - \langle n_1 \rangle^2}}{\langle n_1 \rangle} = \frac{1}{\sqrt{N}} \quad (2.9)$$

This result tells us that when the number of molecules is small (i.e., <100), the cell-to-cell variation is itself a significant fraction of the mean itself.

Such claims are not a mere academic curiosity and have been the subject of careful experimental investigation. On multiple occasions in the remainder of this book, we will examine partitioning data (for mRNA, for proteins, for carboxysomes,

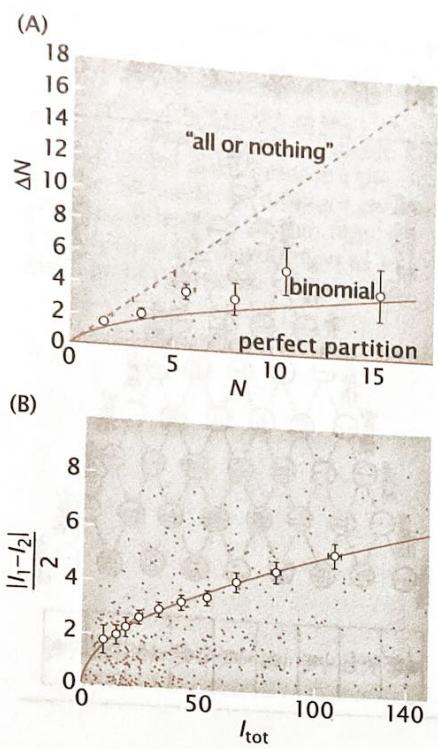


Figure 2.9: Binomial partitioning of mRNA and proteins in *E. coli* during cell division. (A) Difference ΔN in the number of mRNAs between the two daughters given that the mother has N mRNAs. The curves show three possible partitioning mechanisms involving “all or nothing” in which one daughter takes all of the mRNAs, binomial partitioning, and “perfect partitioning” in which each daughter gets exactly half of the proteins from the mother cell. (B) Difference in the fluorescence intensity of the two daughter cells for a fluorescent fusion to a repressor protein as a function of the fluorescence intensity of the mother cell. The line corresponds to binomial partitioning model. (A, adapted from I. Golding et al., *Cell* 123:1025, 2005; B, adapted from N. Rosenfeld et al., *Science* 307:1962, 2005.)

etc.) in which cell lineages are followed and the partitioning of the macromolecules or complexes of interest are measured directly. An example of such data for the case of mRNA and proteins is shown in Figure 2.9. As a result of the ability to fluorescently label individual mRNAs as they are produced, it is possible to characterize the number of such mRNAs in the mother cell before division and then to measure how they partition during the division process. Using fusions of fluorescent proteins to repressor molecules, similar measurements were taken in the context of proteins as shown in Figure 2.9(B). In both of these cases, the essential idea is to count the number of molecules in the mother cell and then to count the number in the two daughters and to see how different they are. Similar data is shown in Figure 18.7 (p. 724) for the case of the organelles in cyanobacteria known as carboxysomes, which contain the all-important enzyme Rubisco so critical to carbon fixation during the process of photosynthesis.

Computational Exploration: Counting mRNA and Proteins by Dilution

As will be highlighted throughout this book, many of our key results depend critically upon the number of molecular actors involved in a given process. As shown in Figure 18.7 (p. 724), for example, fluorescence microscopy has made it possible to track the dynamics and localization of key cellular structures such as the carboxysomes. One of the main conclusions of that study was that the partitioning of the carboxysomes is not a random process since the *statistics* of the partitioning of these organelles between the two daughter cells does not follow the binomial distribution, which would be characteristic of random partitioning. The goal of this Computational Exploration is to explore the statistics of random partitioning from the perspective of simulations, with reference not only to the carboxysome example, but also to clever new ways of taking the molecular census. Using these simulations, we will show how the fluctuations in the partitioning of a fluorescent protein between daughter cells is related to the cells’ intensity as schematized in Figure 2.10. Data from this type of experiment are shown in Figure 19.15 (p. 816).

The advent of fluorescent protein fusions now makes it possible to assess the relative quantities of a given protein of interest on the basis of the fluorescence of its fusion partner. However, for the purposes of the strict comparison between theory and experiment that we are often interested in, it is necessary to have a calibration that permits us to convert between the arbitrary fluorescence units as measured in a microscope and the more interesting molecular counts necessary for plugging into our various formulae or that we use in dissecting some molecular mechanism. To effect this calibration, we exploit the idea of a strict linear relation between the observed intensity and the number of fusion proteins as indicated by the relation

$$I_{tot} = \alpha N_{tot}, \quad (2.10)$$

where α is the unknown calibration factor linking fluorescence and the number of fluorescent molecules (N_{tot}).

As indicated in Figure 2.10, a series of recent experiments all using the same fluctuation method make it possible to

"typical" eukaryotic cells in the remainder of the book. Fibroblasts are associated with animal connective tissue and are notable for secreting the macromolecules of the extracellular matrix.

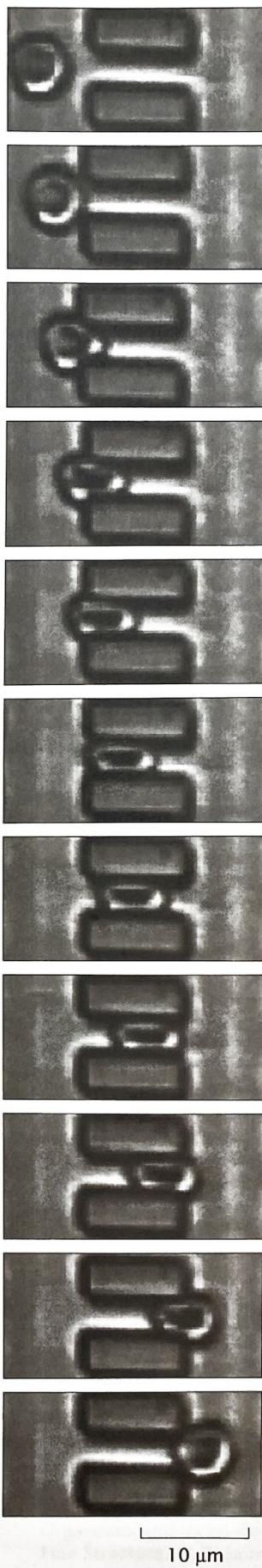
Cells in multicellular organisms can be even more exotic. For example, nerve cells (Figure 2.16F) and rod cells (Figure 2.16G) reveal a great deal more complexity than the examples highlighted above. In these cases, the cell shape is intimately related to its function. In the case of nerve cells, their sinewy appearance is tied to the fact that the various branches (also called "processes") known as dendrites and axons convey electrical signals that permit communication between distant parts of an animal's nervous system. Despite having nuclei with typical eukaryotic dimensions, the cells themselves can extend processes with characteristic lengths of up to tens of centimeters. The structural complexity of rod cells is tied to their primary function of light detection in the retina of the eye. These cells are highly specialized to perform transduction of light energy into chemical energy that can be used to communicate with other cells in the body and, in particular, with brain cells that permit us to be conscious of perceiving images. Rod cells accomplish this task using large stacks of membranes which are the antennas participating in light detection. Figure 2.16 only scratches the surface of the range of cellular size and shape, but at least conveys an impression of cell sizes relative to our standard ruler.

2.2.2 The Cellular Interior: Organelles

As we descend from the scale of the cell itself, a host of new structures known as organelles come into view. The presence of these membrane-bound organelles is one of the defining characteristics that distinguishes eukaryotes from bacteria and archaea. Figure 2.23 shows a schematic of a eukaryotic cell and an associated electron microscopy image revealing some of the key organelles. These organelles serve as the specialized apparatus of cell function in capacities ranging from genome management (the nucleus) to energy generation (mitochondria and chloroplasts) to protein synthesis and modification (endoplasmic reticulum and Golgi apparatus) and beyond. The compartments that are bounded by organellar membranes can have completely different protein and ion compositions. In addition, the membranes of each of these different membrane systems are characterized by distinct lipid and protein compositions.

A characteristic feature of many organelles is that they are compartmentalized structures that are separated from the rest of the cell by membranes. The nucleus is one of the most familiar examples since it is often easily visible using standard light microscopy. If we use

Figure 2.21: Deformability of red blood cells. To measure the deformability of human red blood cells, an array of blocks was fabricated in silicon, each block was $4\text{ }\mu\text{m} \times 4\text{ }\mu\text{m} \times 12\text{ }\mu\text{m}$. The blocks were spaced by $4\text{ }\mu\text{m}$ in one direction and $13\text{ }\mu\text{m}$ in the other. A glass coverslip covered the top of this array of blocks. A dilute suspension of red blood cells in a saline buffer was introduced to the system. A slight pressure applied at one end of the array of blocks provided bulk liquid flow, from left to right in the figure. This liquid flow carried the red blood cells through the narrow passages. Video microscopy captured the results. The figure shows consecutive video fields with the total elapsed time just over one-third of a second. (Courtesy of J. Brody.)



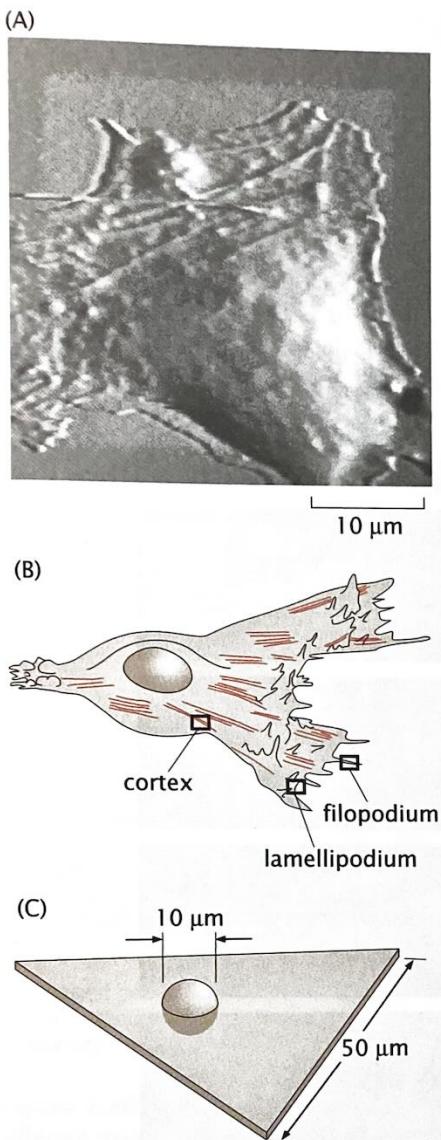


Figure 2.22: Structure of a fibroblast. (A) Atomic-force microscopy image of a fibroblast, (B) cartoon of the external morphology of a fibroblast, and (C) characteristic dimensions of a “typical” fibroblast. (A, adapted from M. Radmacher, *Meth. Cell Biol.* 83:347, 2007.)

the fibroblast as an example, then the cell itself has dimensions of roughly 50 μm , while the nucleus has a characteristic linear dimension of roughly 10 μm , as shown schematically in Figure 2.22. From a functional perspective, the nucleus is much more complex than simply serving as a storehouse for the genetic material. Chromosomes are organized within the nucleus, forming specific domains as will be discussed in more detail in Chapter 8. Transcription as well as several kinds of RNA processing occur in the nucleus. There is a flux of molecules such as transcription factors moving in and completed RNA molecules moving out through elaborate gateways in the nuclear membrane known as nuclear pores. Portions of the genome involved in synthesis of ribosomal RNA are clustered together, forming striking spots that can be seen in the light microscope and are called nucleoli.

Moving outward from the nucleus, the next membranous organelle we encounter is often the endoplasmic reticulum. Indeed, the membrane of the endoplasmic reticulum is contiguous with the membrane of the nuclear envelope. In some cells such as the pancreatic cell shown in Figure 2.24, the endoplasmic reticulum takes up the bulk of the cell interior. This elaborate organelle is the site of lipid synthesis and also the site of synthesis of proteins that are destined to be secreted or incorporated into membranes. It is clear from images such as those in Figures 2.24 and 2.25 that the endoplasmic reticulum can assume different geometries in different cell types and under different conditions. How much total membrane area is taken up by the endoplasmic reticulum? How strongly does the specific membrane morphology affect the total size of the organelle?

Estimate: Membrane Area of the Endoplasmic Reticulum

One of the most compelling structural features of the endoplasmic reticulum is its enormous surface area. To estimate the area associated with the endoplasmic reticulum, we take our cue from Figure 2.24, which suggests that we think of the endoplasmic reticulum as a series of concentric spheres centered about the nucleus. We follow Fawcett (1966), who characterizes the endoplasmic reticulum as forming “lamellar systems of flat cavities, rather uniformly spaced and parallel to one another” as shown in Figure 2.24.

An estimate can be made by adding up the areas from each of the concentric spheres making up our model endoplasmic reticulum. This can be done by simply noticing that the volume enclosed by the endoplasmic reticulum can be written as

$$V_{\text{ER}} = \sum_i A_i d, \quad (2.17)$$

where A_i is the area of the i th concentric sphere and d is the distance between adjacent cisternae. Since two membranes bound each cisterna, the total area of the endoplasmic reticulum membrane is $A_{\text{ER}} = 2 \times \sum_i A_i$. In our model, the total volume of the endoplasmic reticulum can be written as the difference between the volume taken up by the outermost sphere and the volume of the innermost concentric sphere (which is the same as the volume of the nucleus). This results in

$$V_{\text{ER}} = \frac{4\pi}{3} R_{\text{out}}^3 - \frac{4\pi}{3} R_{\text{nucleus}}^3. \quad (2.18)$$



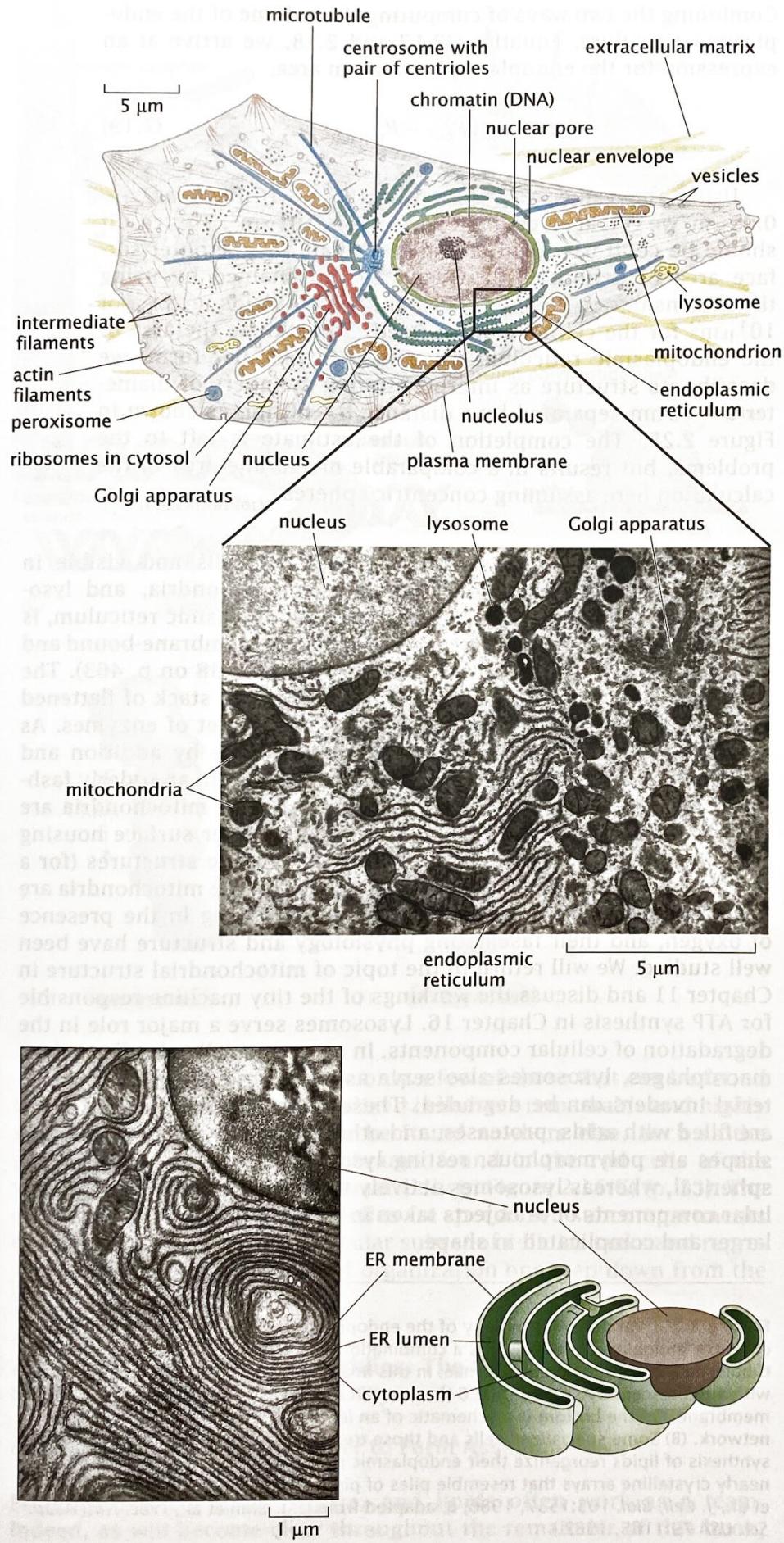


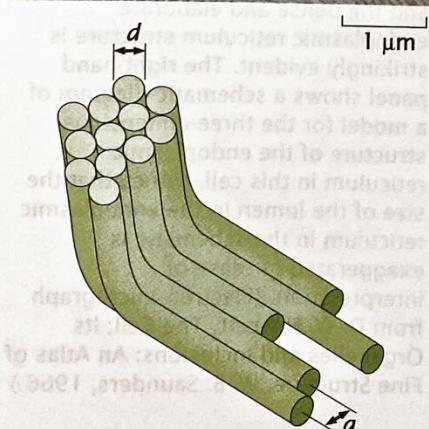
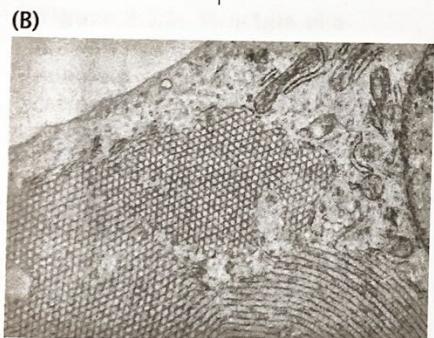
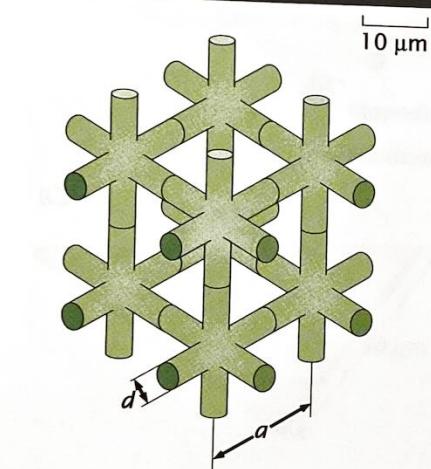
Figure 2.23: Eukaryotic cell and its organelles. The schematic shows a eukaryotic cell and a variety of membrane-bound organelles. A thin-section electron microscopy image shows a portion of a rat liver cell approximately equivalent to the boxed area on the schematic. A portion of the nucleus can be seen in the upper left-hand corner. The most prominent organelles visible in the image are mitochondria, lysosomes, the rough endoplasmic reticulum, and the Golgi apparatus. (Eukaryotic cell from Alberts et al., Molecular Biology of the Cell, 5th ed., New York, Garland Science, 2008; electron micrograph from D. W. Fawcett, *The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure*. W. B. Saunders, 1966.)



Figure 2.24: Electron micrograph and associated schematic of the endoplasmic reticulum (ER). The left-hand panel shows a thin-section electron micrograph of an acinar cell from the pancreas of a bat. The nucleus is visible at the upper right and the dense and elaborate endoplasmic reticulum structure is strikingly evident. The right-hand panel shows a schematic diagram of a model for the three-dimensional structure of the endoplasmic reticulum in this cell. Notice that the size of the lumen in the endoplasmic reticulum in the schematic is exaggerated for ease of interpretation. (Electron micrograph from D. W. Fawcett, *The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure*. W. B. Saunders, 1966.)

Combining the two ways of computing the volume of the endoplasmic reticulum, Equations 2.17 and 2.18, we arrive at an expression for the endoplasmic reticulum area,

$$A_{ER} = \frac{8\pi}{3d} (R_{out}^3 - R_{nucleus}^3). \quad (2.19)$$



Using the values $R_{nucleus} = 5 \mu\text{m}$, $R_{out} = 10 \mu\text{m}$, and $d = 0.05 \mu\text{m}$, we get, at an estimate, $A_{ER} = 15 \times 10^4 \mu\text{m}^2$. This result should be contrasted with a crude estimate for the outer surface area of a fibroblast, which can be obtained by using the dimensions in Figure 2.22(C) and which yields an area of $10^4 \mu\text{m}^2$ for the cell membrane itself. To estimate the area of the endoplasmic reticulum when it is in reticular form, we describe its structure as interpenetrating cylinders of diameter $d \approx 10 \text{ nm}$ separated by a distance $a \approx 60 \text{ nm}$, as shown in Figure 2.25. The completion of the estimate is left to the problems, but results in a comparable membrane area to the calculation here assuming concentric spheres.

The other major organelles found in most cells and visible in Figure 2.23 include the Golgi apparatus, mitochondria, and lysosomes. The Golgi apparatus, similar to the endoplasmic reticulum, is largely involved in processing and trafficking of membrane-bound and secreted proteins (for another view, see Figure 11.38 on p. 463). The Golgi apparatus is typically seen as a pancake-like stack of flattened compartments, each of which contains a distinct set of enzymes. As proteins are processed for secretion, for example by addition and remodeling of attached sugars, they appear to pass in an orderly fashion through each element in the Golgi stack. The mitochondria are particularly striking organelles with a smooth outer surface housing an elaborately folded system of internal membrane structures (for a more detailed view, see Figure 11.39 on p. 464). The mitochondria are the primary site of ATP synthesis for cells growing in the presence of oxygen, and their fascinating physiology and structure have been well studied. We will return to the topic of mitochondrial structure in Chapter 11 and discuss the workings of the tiny machine responsible for ATP synthesis in Chapter 16. Lysosomes serve a major role in the degradation of cellular components. In some specialized cells such as macrophages, lysosomes also serve as the compartment where bacterial invaders can be degraded. These membrane-bound organelles are filled with acids, proteases, and other degradative enzymes. Their shapes are polymorphous; resting lysosomes are simple and nearly spherical, whereas lysosomes actively involved in degradation of cellular components or of objects taken in from the outside may be much larger and complicated in shape.

Figure 2.25: Variable morphology of the endoplasmic reticulum. (A) In most cultured cells, the endoplasmic reticulum is a combination of a web-like reticular network of tubules and larger flattened cisternae. In this image, a cultured fibroblast was stained with a fluorescent dye called DiOC6 that specifically labels endoplasmic reticulum membrane. On the bottom is a schematic of an idealized three-dimensional reticular network. (B) Some specialized cells and those treated with drugs that regulate the synthesis of lipids reorganize their endoplasmic reticulum to form tightly-packed, nearly crystalline arrays that resemble piles of pipes. (A, adapted from M. Terasaki et al., *J. Cell. Biol.* 103:1557, 1986; B, adapted from D. J. Chin et al., *Proc. Natl Acad. Sci. USA* 79:1185, 1982.)

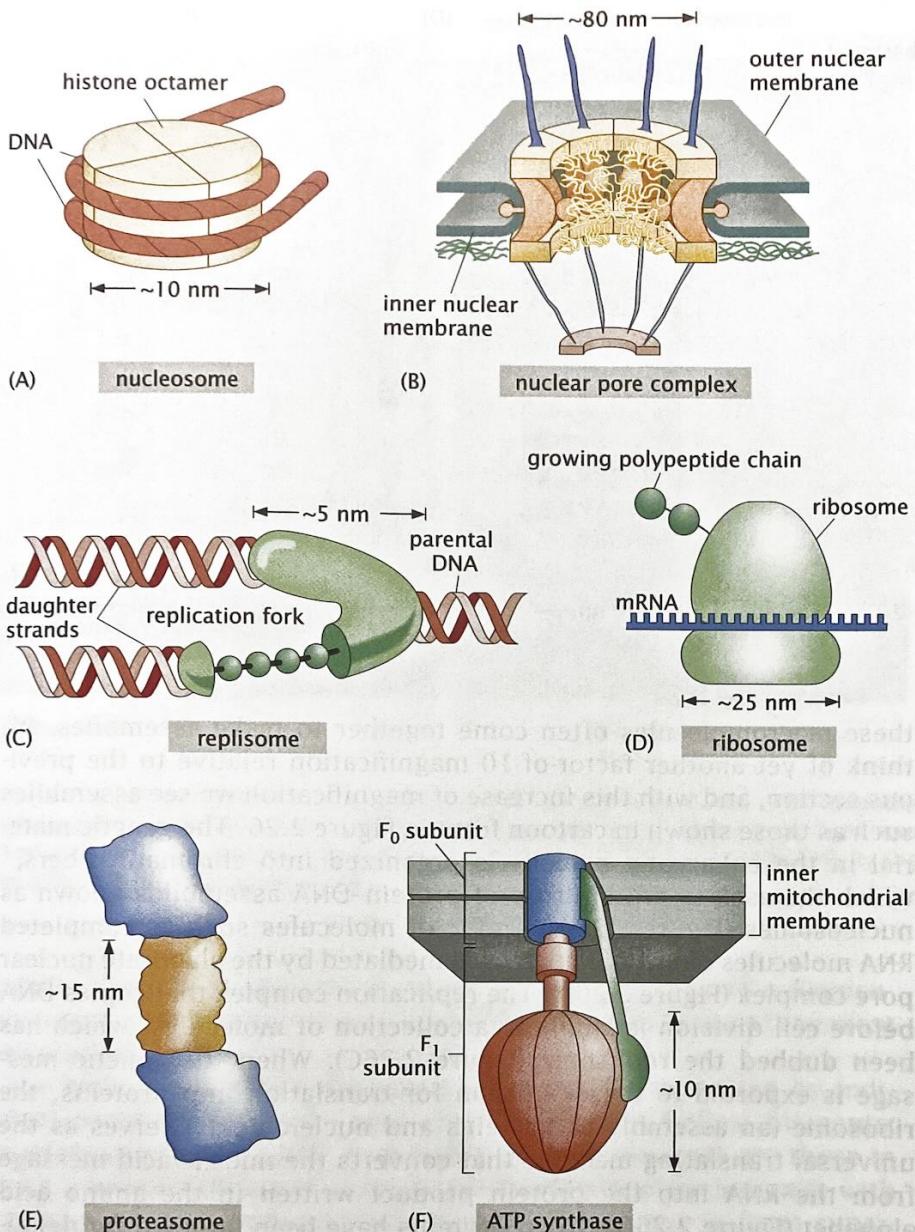


Figure 2.26: The macromolecular assemblies of the cell. (A) The nucleosome is a complex of eight protein units and DNA. (B) The nuclear pore complex mediates the transport of material in and out of the nucleus. (C) The replisome is a complex of proteins that mediate the copying of the genetic material. (D) The ribosome reads mRNA and synthesizes the corresponding polypeptide chain. (E) The proteasome degrades proteins. (F) ATP synthase is a large complex that synthesizes new ATP molecules.

These common organelles are only a few of those that can be found in eukaryotic cells. Some specialized cells have remarkable and highly specialized organelles that can be found nowhere else, such as the stacks of photoreceptive membranes found in the rod cells of the visual system as indicated schematically in Figure 2.16(G) (p. 53). The common theme is that all organelles are specialized subcompartments of the cell that perform a particular subset of cellular tasks and represent a smaller, discrete layer of organization one step down from the whole cell.

2.2.3 Macromolecular Assemblies: The Whole is Greater than the Sum of the Parts

Macromolecules Come Together to Form Assemblies

Proteins, nucleic acids, sugars, and lipids often work as a team! Indeed, as will become clear throughout the remainder of this book,

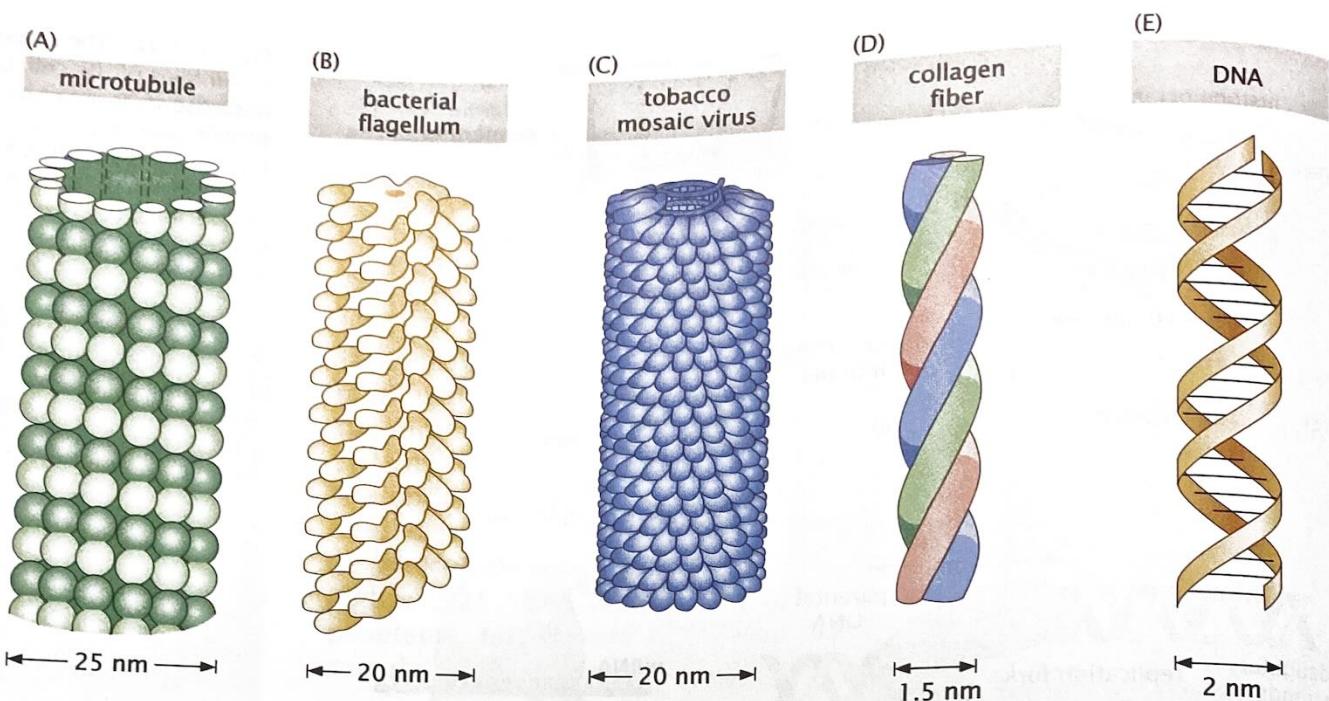


Figure 2.27: Helical motifs of molecular assemblies. Macromolecular assemblies have a variety of different helical structures, some formed from individual monomeric units (such as (A)–(C)), others resulting from coils of proteins (D), and yet others made up of paired nucleotides (E).

these macromolecules often come together to make assemblies. We think of yet another factor-of-10 magnification relative to the previous section, and with this increase of magnification we see assemblies such as those shown in cartoon form in Figure 2.26. The genetic material in the eukaryotic nucleus is organized into chromatin fibers, which themselves are built up of protein-DNA assemblies known as nucleosomes (Figure 2.26A). Traffic of molecules such as completed RNA molecules out of the nucleus is mediated by the elaborate nuclear pore complex (Figure 2.26B). The replication complex that copies DNA before cell division is similarly a collection of molecules, which has been dubbed the replisome (Figure 2.26C). When the genetic message is exported to the cytoplasm for translation into proteins, the ribosome (an assembly of proteins and nucleic acids) serves as the universal translating machine that converts the nucleic acid message from the RNA into the protein product written in the amino acid alphabet (Figure 2.26D). When proteins have been targeted for degradation, they are sent to another macromolecular assembly known as the proteasome (Figure 2.26E). The production of ATP in mitochondria is similarly mediated by a macromolecular complex known as ATP synthase (Figure 2.26F). The key idea of this subsection is to show that there is a very important level of structure in cells that is built around complexes of individual macromolecules and with a characteristic length scale of 10 nm.

Helical Motifs Are Seen Repeatedly in Molecular Assemblies

A second class of macromolecular assemblies, characterized not by function but rather by structure, is the wide variety of helical macromolecular complexes. Several representative examples are depicted in Figure 2.27. In Figure 2.27(A), we show the geometric structure of microtubules. As will be described in more detail later, these structures are built up of individual protein units called tubulin. A second example shown in Figure 2.27(B) is the bacterial flagellum of *E. coli*. Here too, the same basic structural idea is repeated, with the helical geometry built up from individual protein units, in this case

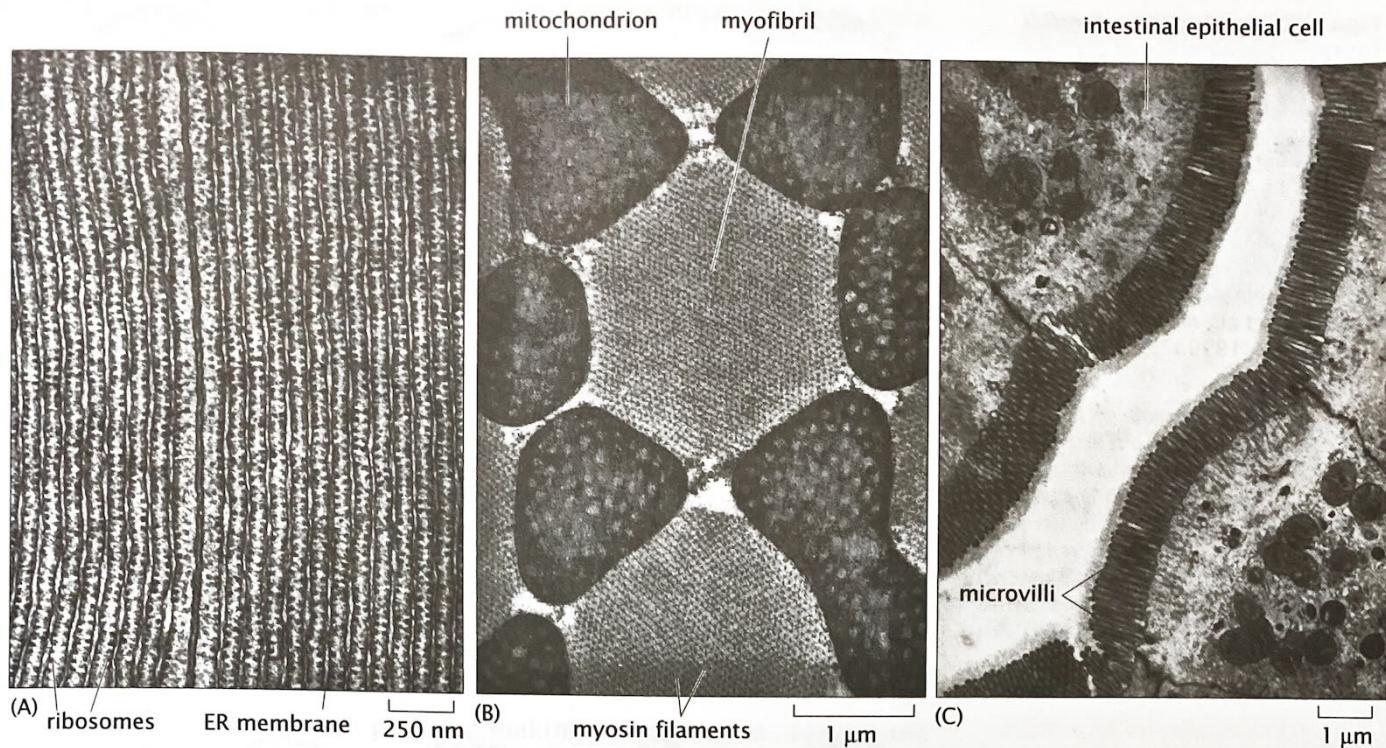


Figure 2.28: Ordered macromolecular assemblies. Collage of examples of macromolecules organized into superstructures. (A) Ribosomes on the endoplasmic reticulum ("rough ER"), (B) myofibrils in the flight muscle, and (C) microvilli at the epithelial surface. (A–C, adapted from D. W. Fawcett, *The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure*. W. B. Saunders, 1966.)

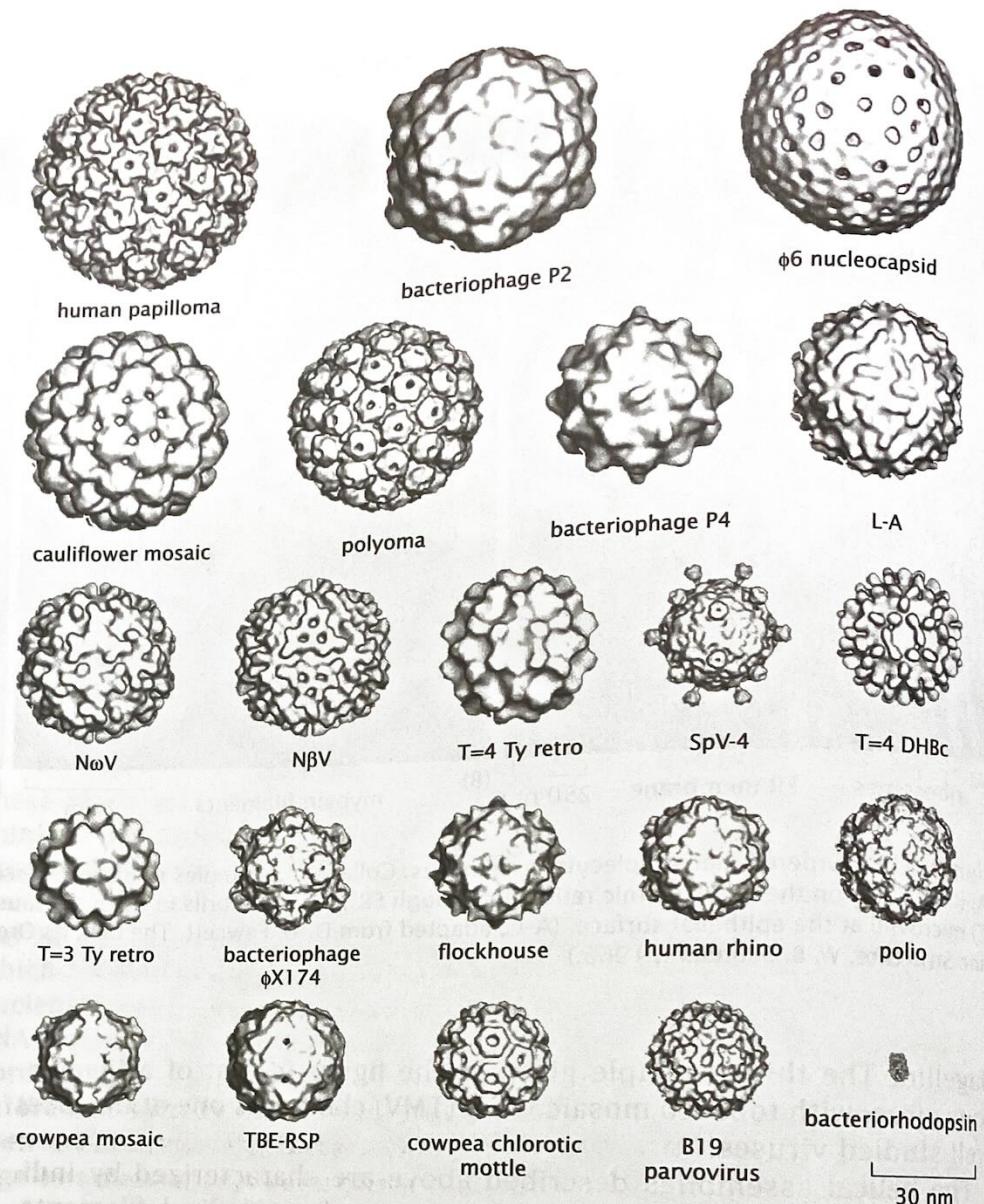
flagellin. The third example given in the figure is that of a filamentous virus, with tobacco mosaic virus (TMV) chosen as one of the most well-studied viruses.

The helical assemblies described above are characterized by individual protein units that come together to form helical filaments. An alternative and equally remarkable class of filaments are those in which α -helices (chains of amino acids forming protein subunits with a precise, helical geometry) wind around each other to form superhelices. The particular case study that will be of most interest in subsequent discussions is that of collagen, which serves as one of the key components in the extracellular matrix of connective tissues and is one of the major protein products of the fibroblast cells introduced earlier in the chapter (see Figure 2.16 on p. 53). The final example is the iconic figure of the DNA molecule itself.

Macromolecular Assemblies Are Arranged in Superstructures

Assemblies of macromolecules can interact with each other to create striking instances of cellular hardware with a size comparable to organelles themselves. Figure 2.28 depicts several examples. Figure 2.28(A) shows the way in which ribosomes are organized on the endoplasmic reticulum with a characteristic spacing that is comparable to the size of the ribosomes (≈ 20 nm). A second stunning example is the organization of myofibrils in muscles as shown in Figure 2.28(B). This figure shows the juxtaposition of the myofibrils and mitochondria. The myofibrils themselves are an ordered arrangement of actin filaments and myosin motors, as will be discussed in more detail in

Figure 2.29: Structures of viral capsids. The regularity of the structure of viruses has enabled detailed, atomic-level analysis of their construction patterns. This gallery shows a variety of the different geometries explored by the class of nearly spherical viruses. For size comparison, a large protein, bacteriorhodopsin, is shown in the bottom right. (Adapted from T. S. Baker et al., *Microbiol. Mol. Biol. Rev.* 63:862, 1999.)



Chapter 16. The last example, shown in Figure 2.28(C), is of the protrusions of microvilli at the surface of an epithelial cell. These microvilli are the result of collections of parallel actin filaments. The list of examples of orchestration of collections of macromolecules could go on and on and should serve as a reminder of the many different levels of structural organization found in cells.

2.2.4 Viruses as Assemblies

Viruses are one of the most impressive and beautiful examples of macromolecular assembly. These assemblies are a collection of proteins and nucleic acids (though many viruses have lipid envelopes as well) that form highly ordered and symmetrical objects with characteristic sizes of tens to hundreds of nanometers. The architecture of viruses is usually a protein shell where the so-called capsid is made up of a repetitive packing of the same protein subunits over and over to form an icosahedron. Within the capsid, the virus packs its genetic material, which can be either single-stranded or double-stranded DNA or RNA depending upon the type of virus. Figure 2.29 is a gallery of the capsids of a number of different viruses. Different viruses have different elaborations on this basic structure and can include lipid coats,

surface receptors, and internal molecular machines such as polymerases and proteases. One of the most amazing features of viruses is that by hijacking the machinery of the host cell, the viral genome commands the construction of its own inventory of parts within the host cell and then, in the crowded environment of that host, assembles into infectious agents prepared to repeat the life cycle elsewhere.

Human immunodeficiency virus (HIV) is one of the viruses that has garnered the most attention in recent years. Figure 2.30 shows cryo-electron microscopy images of mature HIV virions and gives a sense of both their overall size (roughly 130 nm) and their internal structure. In particular, note the presence of an internal capsid shaped like an ice-cream cone. This internal structure houses the roughly 10 kilobase (kb) RNA viral genome. As with our analysis of the inventory of a cell considered earlier in the chapter, part of developing a “feeling for the organism” is to get a sense of the types and numbers of the different molecules that make up that organism. In the case of HIV, these numbers are interesting for many reasons, including that they say something about the “investment” that the infected cell has to make in order to construct new virions.

For our census of an HIV virion, we need to examine the assembly of the virus. In particular, one of the key products of its roughly 10 kb genome is a polyprotein, a single polypeptide chain containing what will ultimately be distinct proteins making up the capsid, matrix, and nucleocapsid, known as Gag and shown schematically in Figure 2.31. The formation of the *immature* virus occurs through the association of

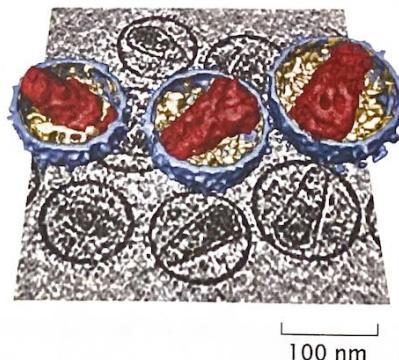


Figure 2.30: Structure of HIV viruses. The planar image shows a single frame from an electron microscopy tilt series. The three-dimensional images show reconstructions of the mature viruses featuring the ice-cream-cone-shaped capsid on the interior. (Adapted from J. A. G. Briggs et al., *Structure* 14:15, 2006.)

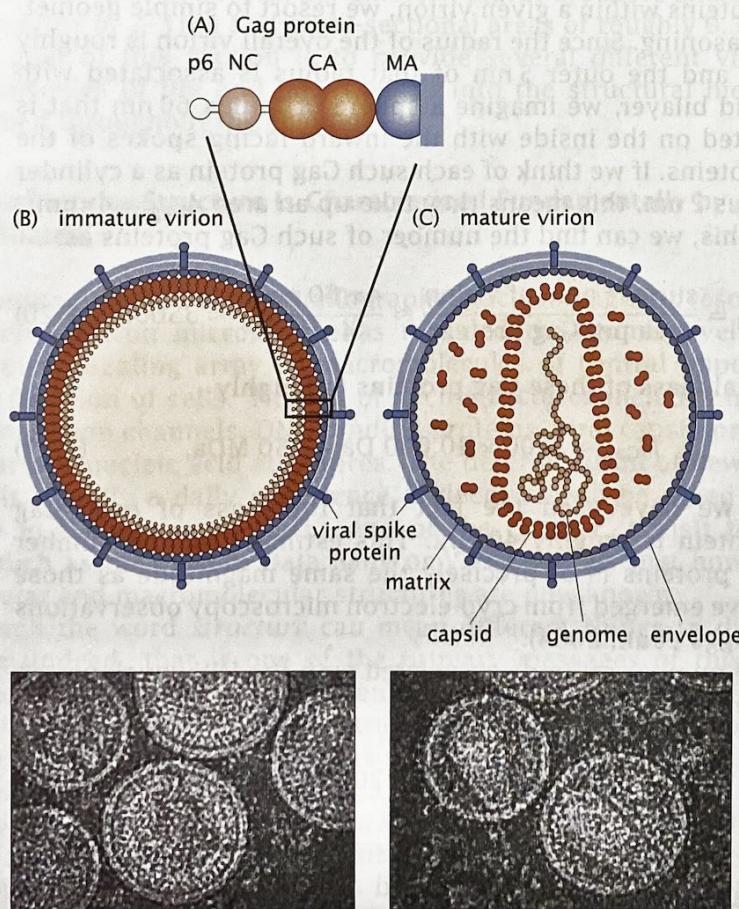


Figure 2.31: HIV architecture. (A) Schematic of the Gag polyprotein, a 41,000 Da architectural building block. (B) Immature virions showing the lipid bilayer coat and the uncut Gag shell on the interior. (C) Mature virions in which the Gag protein has been cut by proteases and the separate components have assumed their architectural roles in the virus. The associated electron microscopy images show actual data for each of the cartoons. (Adapted from J. A. G. Briggs et al., *Nat. Struct. Mol. Biol.* 11:672, 2004.)

the N-terminal ends of these Gag proteins with the lipid bilayer of the host cell, with the C-termini pointing radially inward like the spokes of a three-dimensional wheel. (N-terminus and C-terminus refer to the two structurally distinct ends of the polypeptide chain. During protein synthesis, translation starts at the N-terminus and finishes at the C-terminus.) As more of these proteins associate on the cell surface, the nascent virus begins to form a bud on the cell surface, ultimately resulting in spherical structures like those shown in Figure 2.31. During the process of viral maturation, a viral protease (an enzyme that cuts proteins) clips the Gag protein into its component pieces known as matrix (MA), capsid (CA), nucleocapsid (NC), and p6. The matrix forms a shell of proteins just inside of the lipid bilayer coat. The capsid proteins form the ice-cream-cone-shaped object that houses the genetic material and the nucleocapsid protein is complexed with the viral RNA.

different types of lipids forming the viral envelope (see Brügger et al., 2006). Interestingly, the diversity of lipids in the HIV envelope is enormous, with the lipid composition of the viral envelope distinct from that of the host cell membrane. The measured total number of different lipids is roughly 300,000. Further analysis of the parts list of HIV is left to the problems at the end of the chapter.

Ultimately, viruses are one of the most interesting cases of the power of macromolecular assembly. These intriguing machines occupy a fuzzy zone at the interface between the living and the nonliving.

2.2.5 The Molecular Architecture of Cells: From Protein Data Bank (PDB) Files to Ribbon Diagrams

If we continue with another factor of 10 in our powers-of-10 descent, we find the individual macromolecules of the cell. In particular, this increase in spatial resolution reveals four broad categories of macromolecules: lipids, carbohydrates, nucleic acids, and proteins. As was shown in Chapter 1 (p. 3), these four classes of molecules make up the stuff of life and have central status in building cells both architecturally and functionally. Though often these molecules are highly anisotropic (for example, a DNA molecule is usually many orders of magnitude longer than it is wide), their characteristic scale is between 1 and 10 nm. For example, as shown earlier in the chapter, a “typical” protein has a size of several nanometers. Lipids are more anisotropic, with lengths of 2–3 nm and cross-sectional areas of roughly 0.5 nm^2 .

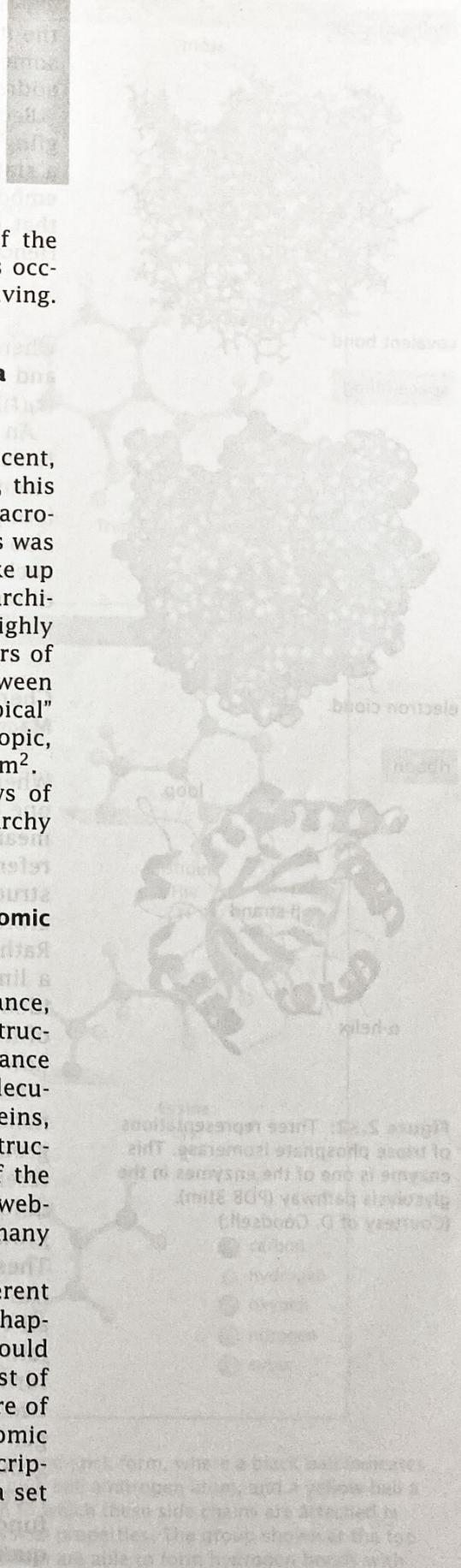
The goal of this section is to provide several different views of the molecules of life and how they fit into the structural hierarchy described throughout the chapter.

Macromolecular Structure Is Characterized Fundamentally by Atomic Coordinates

The conjunction of X-ray crystallography, nuclear magnetic resonance, and cryo-electron microscopy has revealed the atomic-level structures of a dazzling array of macromolecules of central importance to the function of cells. The list of such structures includes molecular motors, ion channels, DNA-binding proteins, viral capsid proteins, and various nucleic acid structures. The determination of new structures is literally a daily experience. Indeed, as will be asked of the reader in the problems at the end of the chapter, a visit to websites such as the Protein Data Bank or VIPER reveals just how many molecular and macromolecular structures are now known.

Though the word *structure* can mean different things to different people (indeed, that is one of the primary messages of this chapter and Chapter 8), most self-identified structural biologists would claim that the determination of structure ultimately refers to a list of atomic coordinates for the various atoms making up the structure of interest. As an example, Figure 1.3 (p. 6) introduced detailed atomic portraits of nucleic acids, proteins, lipids, and sugars. In such descriptions, the structural characterization of the system amounts to a set of coordinates

$$\mathbf{r}_i = x_i \mathbf{i} + y_i \mathbf{j} + z_i \mathbf{k}, \quad (2.23)$$



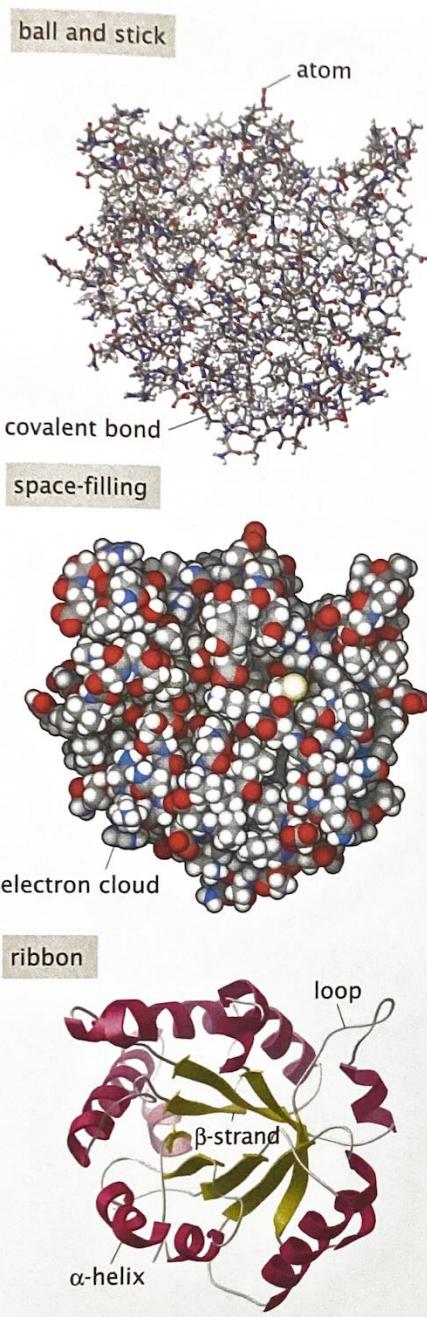


Figure 2.32: Three representations of triose phosphate isomerase. This enzyme is one of the enzymes in the glycolysis pathway (PDB 3tim). (Courtesy of D. Goodsell.)

where, having chosen some origin of coordinates, the coordinates of the i th atom in the structure are given by (x_i, y_i, z_i) . That is, we have some origin of Cartesian coordinates and every atomic position is an address on this three-dimensional grid.

Because the macromolecules of the cell are subject to incessant jiggling due to collisions with each other and the surrounding water, a static picture of structure is incomplete. The structural snapshots embodied in atomic coordinates for a given structure miss the fact that each and every atom is engaged in a constant thermal dance. Hence, the coordinates of Equation 2.23 are really of the form

$$\mathbf{r}_i(t) = x_i(t)\mathbf{i} + y_i(t)\mathbf{j} + z_i(t)\mathbf{k}, \quad (2.24)$$

where the t reminds us that the coordinates depend upon time and what is measured in experiments might be best represented as $\langle \mathbf{r}_i(t) \rangle_{\text{time}}$, where the brackets $\langle \rangle_{\text{time}}$ signify an average over time.

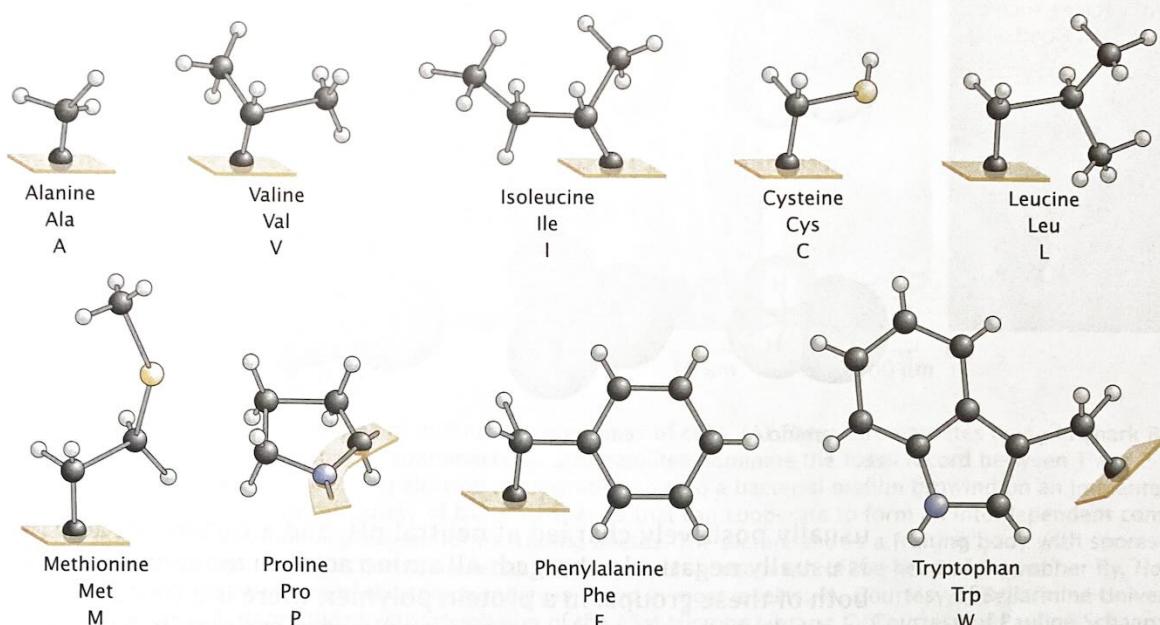
An example of an atomic-level representation of one of the key proteins of the glycolysis pathway is shown in Figure 2.32. We choose this example because glycolysis will arise repeatedly throughout the book (see p. 191 for example) as a canonical metabolic pathway. The figure also shows several alternative schemes for capturing these structures such as using ribbon diagrams which highlight the ways in which the different amino acids come together to form elements of secondary structure such as α -helices and β -sheets.

Chemical Groups Allow Us to Classify Parts of the Structure of Macromolecules

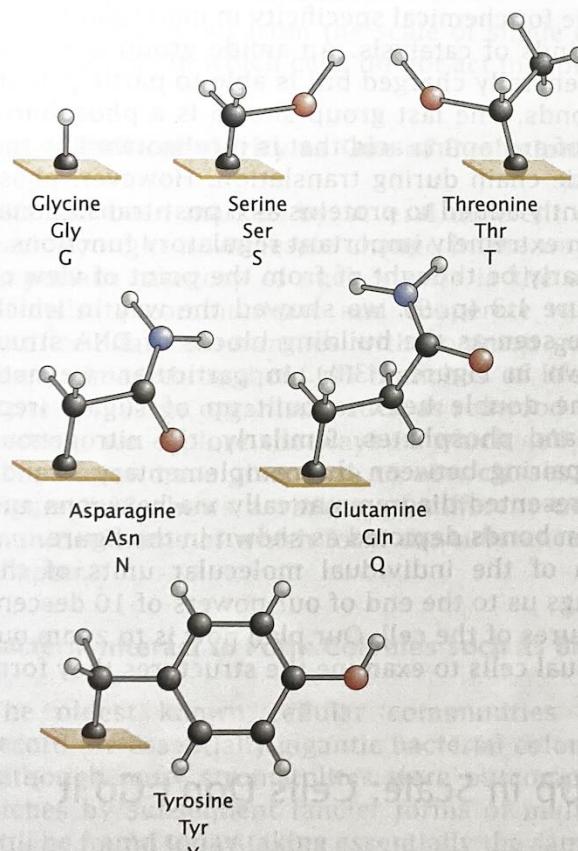
When thinking about the structures of the macromolecules of the cell, one of the most important ways to give those structures functional meaning (as opposed to just a collection of coordinates) is through reference to the chemical groups that make them up. For example, the structure of the protein shown in Figures 1.3 and 2.32 is not just an arbitrary arrangement of carbons, nitrogens, oxygens, and hydrogens. Rather, this structure reflects the fact that the protein is made up of a linear sequence of amino acids that each have their own distinct identity as shown in Figure 2.33. The physical and chemical properties of these amino acids dictate the folded shape of the protein as well as how it functions.

Amino acids are but one example of a broader class of sub-nanometer-scale structural building blocks known as “chemical groups.” These chemical groups occur with great frequency in different macromolecules and, like the amino acids, each has its own unique chemical identity. Figure 2.34 shows a variety of chemical groups that are of interest in biochemistry and molecular biology. These are all biologically important chemical functional groups that can be attached to a carbon atom as shown in Figure 2.34 and are all found in protein structures. The methyl and phenyl groups contain only carbon and hydrogen and are hence hydrophobic (unable to form hydrogen bonds with water). To the right of these are shown two chemically similar groups, alcohol and thiol, consisting of oxygen or sulfur plus a single hydrogen. The key feature of these two groups is that they are highly reactive and can participate in chemical reactions forming new covalent bonds. Amino acids containing these groups are frequently important active site residues in enzyme-catalyzed reactions. The next row starts with a nitrogen-containing amino group, which is

HYDROPHOBIC



POLAR



CHARGED

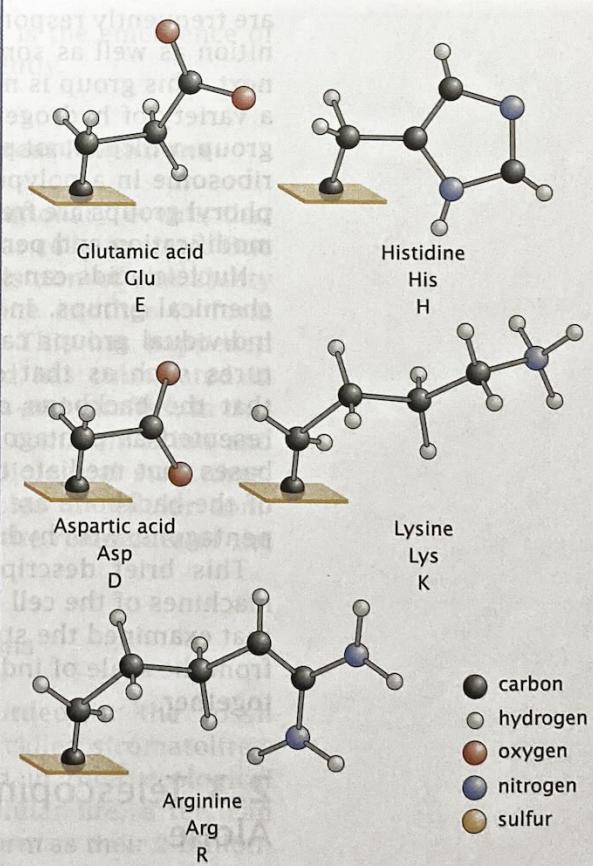
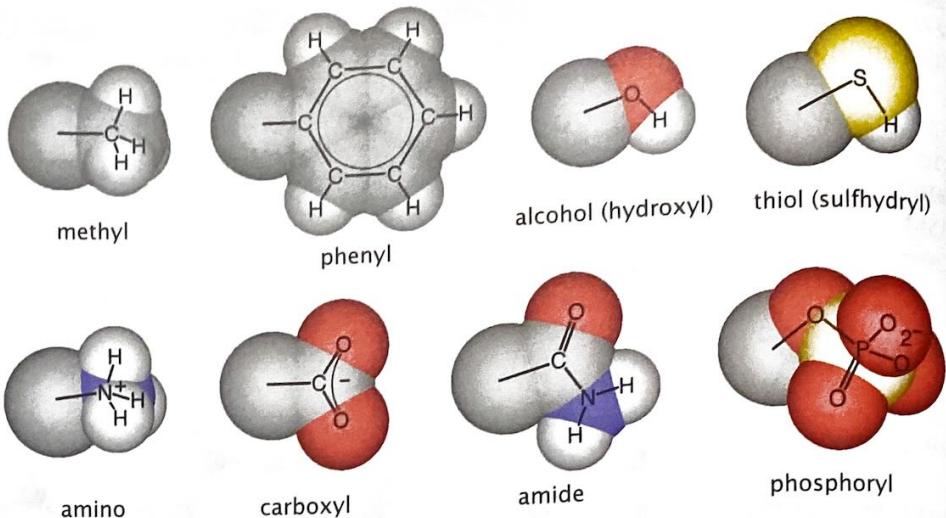


Figure 2.33: Amino acid side chains. The amino acids are represented here in ball-and-stick form, where a black ball indicates a carbon atom, a small white ball a hydrogen atom, a red ball an oxygen atom, a blue ball a nitrogen atom, and a yellow ball a sulfur atom. Only the side chains are shown. The peptide backbone of the protein to which these side chains are attached is indicated by an orange tile. The amino acids are subdivided based upon their physical properties. The group shown at the top are hydrophobic and tend to be found on the interior of proteins. Those at the bottom are able to form hydrogen bonds with water and are often found on protein surfaces.

Figure 2.34: Chemical groups. These are some of the most common groups found in organic molecules such as proteins. (Courtesy of D. Goodsell.)



usually positively charged at neutral pH, and a carboxylic acid, which is usually negatively charged. All amino acids in monomeric form have both of these groups. In a protein polymer, there is a free amino group at the N-terminus of the protein and a free carboxylic acid group at the C-terminus of the protein. Several amino acids also contain these groups as part of their side chains and the charge-based interactions are frequently responsible for chemical specificity in molecular recognition as well as some kinds of catalysis. An amide group is shown next. This group is not generally charged but is able to participate in a variety of hydrogen bonds. The last group shown is a phosphoryl group, which is not part of any amino acid that is incorporated by the ribosome in a polypeptide chain during translation. However, phosphoryl groups are frequently added to proteins as a post-translational modification and perform extremely important regulatory functions.

Nucleic acids can similarly be thought of from the point of view of chemical groups. In Figure 1.3 (p. 9), we showed the way in which individual groups can be seen as the building blocks of DNA structures such as that shown in Figure 1.3(A). In particular, we note that the backbone of the double helix is built up of sugars (represented as pentagons) and phosphates. Similarly, the nitrogenous bases that mediate the pairing between the complementary strands of the backbone are represented diagrammatically via hexagons and pentagons, with hydrogen bonds depicted as shown in the figure.

This brief description of the individual molecular units of the machines of the cell brings us to the end of our powers-of-10 descent that examined the structures of the cell. Our plan now is to zoom out from the scale of individual cells to examine the structures they form together.