Chapter 2

What and Where: Construction Plans for Cells and Organisms

"Although not everyone is mindful of it all cell biologists have two cells of interest: the one they are studying and *Escherichia coli*." - F. Neidhardt

Chapter Overview: In Which We Consider the Size of Cells and the Nature of Their Contents

Cells come in a dazzling variety of shapes and sizes. Even so, their molecular inventories share many common features, reflecting the underlying biochemical unity of life. In this chapter, we introduce the bacterium *Escherichia coli* (we will abbreviate this cell type as *E. coli* throughout the book) as our biological standard ruler. This cell serves as the basis for a first examination of the inventory of cells and will permit us to get a sense of the size of cells and the nature of their contents. Indeed, using simple estimates, we will take stock of the genome size, numbers of lipids and proteins and the ribosome content of bacteria. With the understanding revealed by *E. coli* in hand, we then take a powers-of-ten journey down and up from the scale of individual cells. Our downward journey will examine organelles within cells, macromolecular assemblies ranging from ribosomes to viruses and then the macromolecules that are the engines of cellular life. Our upward journey from the scale of individual cells will examine a second class of biological structures, namely those resulting from multicellularity, this time with an emphasis on how cells act together in contexts ranging from bacterial biofilms to the networks of neurons in the brain.

2.1 An Ode to *E. coli*

Scientific observers of the natural world have been intrigued by the processes of life for many thousands of years as evidenced by early written records from Aristotle, for example. Early thinkers wondered about the nature of life and its "indivisible" units in much the same way that they mused about the fundamental units of matter. Just as physical scientists arrived at a consensus that the fundamental unit of matter is the atom (at least for chemical transactions), likewise, observers of living organisms have agreed that the indivisible unit of life is the cell. Nothing smaller than a cell can be shown to be alive in a sense that is generally agreed upon. At the same time, there are no currently known reasons to attribute some higher "living" status to multicellular organisms.

Cells are able to consume energy from their environments and use that energy to create ordered structures. They can also harness energy from the environment to create new cells. A standard definition of life merges the features of metabolism (that is, consumption and use of energy from the environment) and replication (giving offspring that resemble the original organism). Stated simply, the cell is the smallest unit of replication (though viruses are also replicative units, but depend upon their infected host to provide much of the machinery making this replication possible).

The recognition that the cell is the fundamental unit of biological organization originated in the seventeenth century with the microscopic observations of Hooke and van Leeuwenhoek. This idea was put forth as the modern cell theory by Schwann, Schleiden and Virchow in the mid-nineteenth century and was confirmed unequivocally by Pasteur shortly thereafter and repeatedly in the time since. Biologists agree that all forms of life share cells as the basis of their organization. It is also generally agreed that all living organisms on Earth shared a common ancestor several billion years ago that would be recognized as a cell by any modern biologist.

In terms of understanding the basic rules governing metabolism, replication and life more generally, one cell type as the basis of experimental investigations of these mechanisms should be as good as any other. For practical reasons, however, biologists have focused on a few particular types of cell to try to illuminate these general issues. Among these, the human intestinal inhabitant *E. coli* stands unchallenged as the most useful and important representative of the living world in the biologist's laboratory.

Several properties of *E. coli* have contributed to its great utility and has made it a source of repeated discoveries. First, it is easy to isolate because it is present in great abundance in human fecal matter. Unlike most other bacteria that populate the human colon, *E. coli* is able to grow well in the presence of oxygen. In the laboratory, it replicates rapidly and can easily adjust to changes in its environment including changes in nutrients. In addition, it is nearly routine to deliberately manipulate these cells (using insults such as radiation or chemicals, for example) to produce mutants. Mutant organisms are those which differ from their parents and from other members of their species found in the wild because of specific changes in DNA sequence which give rise to biologically significant

differences. For example, *E. coli* is normally able to synthesize purines for DNA and RNA on its own from sugar as a nutrient source. However, particular mutants of *E. coli* with enzymatic deficiencies in these pathways have lost the ability to make their own purines and become reliant on being fed precursors for these molecules. A more familiar and frightening example is the way in which bacteria acquire antibiotic resistance. Throughout the book we will be using specific examples of biological phenomena to illustrate general physical principles that are relevant to life. Often, we will have recourse to *E. coli* because of particular experiments that have been performed on this organism. Further, even when we speak of experiments on other cells or organisms, often *E. coli* will be behind the scenes coloring our thinking.

2.1.1 The Bacterial Standard Ruler

The Bacterium *E. coli* Will Serve as Our Standard Ruler

Throughout the book we will discuss many different cells which all share with *E. coli* the fundamental biological directive to convert energy from the environment into structural order and to perpetuate their species. On Earth, it is observed that there are certain minimal requirements for the perpetuation of cellular life. These are not necessarily absolute physical requirements, but in the competitive environment of our planet, all surviving cells share these features in common. These include a DNA-based genome, mechanisms to transcribe DNA into RNA and subsequently, translation mechanisms using ribosomes to convert information in RNA sequences into protein sequence and protein structure. Within those individual cells, there are many substructures with interesting functions. Larger than the cell there are also structures of biological interest that arise because of cooperative interactions between many cells or even different organisms. In this chapter, we will begin with the cell as the fundamental unit of biological organization using *E. coli* as the standard reference and standard ruler. We will then look at smaller structures within cells and finally, larger multicellular structures, zooming in and out from our fundamental cell reference frame.

Fig. 2.1 shows several experimental pictures of an *E. coli* cell and its schematization into our standard ruler. In particular, the electron micrograph in fig. 2.1 shows that these bacteria have a rod-like morphology with a typical length between 1 and 2 microns and a diameter between 1/2 and 1 micron. To put the standard ruler in perspective, we note that with its characteristic length scale of 1 micron, it would take roughly fifty such cells lined up end to end in order to measure out the width of a human hair. On the other hand, we would need to divide the cell into roughly five hundred slices of equal width in order to measure out the diameter of a DNA molecule.

Note that the average size of these cells depends on the nutrients they are provided, with those growing in richer media also having a larger size. Our reference growth condition throughout the book will be a chemically defined solution referred to by microbiologists as “minimal media” with glucose as the

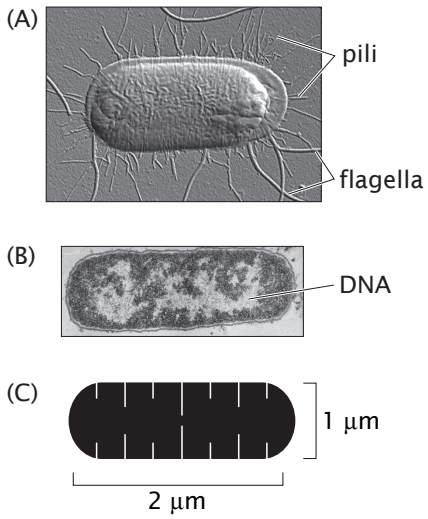

Figure 2.1: *E. coli* as a standard ruler for characterizing spatial scales. (A) Atomic force microscopy image of an *E. coli* cell, (B) electron micrograph of an *E. coli* bacterium, (C) the *E. coli* ruler. (C, courtesy of C. T. Lim.)

sole carbon source. "Minimal media" refers to a completely chemically defined mixture of salts, sugars, amino acids and vitamins that can support the growth of a microorganism. In the laboratory, bacteria are often grown in "rich media", which are poorly defined but nutrient-rich mixtures of extracts from organic materials such as yeast cultures or cow brains. Although microorganisms can grow very rapidly in rich media, they are rarely used for biochemical studies because their exact contents are not known. In minimal media, however, it is easy to simply leave out or add a single compound (for example, a single amino acid such as tryptophan) and measure the effects of that compound on the microorganism's growth.

Because of its central role as the quantitative standard in the remainder of the book, it is useful to further characterize the geometry of *E. coli*. One example in which we will need a better sense of the geometry of cells and their internal compartments is in the context of reconciling *in vitro* (i.e. in test tubes) and *in vivo* (i.e. in living cells) experiments. Results from experiments done *in vitro* are based upon the free concentrations of different molecular species. On the other hand, in *in vivo* situations we might know the number of copies of a given molecule such as a transcription factor (the proteins that regulate expression of genes by binding to DNA). To reconcile these two pictures, we will need the cellular volume to make the translation between molecular counts

and concentrations. Similarly, when examining the distribution of membrane proteins on the cell surface, we will need a sense of the cell area to estimate the mean spacing between these proteins which will tell us about the extent of interaction between them. For most cases of interest in this book, it suffices to attribute a volume $V_{E.coli} \approx 1\ \mu\text{m}^3 = 1$ fL to *E. coli* and an area of roughly $A_{E.coli} \approx 6\ \mu m^2$ (see the problems to actually work out these numbers from known cellular dimensions).

2.1.2 Taking the Molecular Census

In the remainder of this section, we will proceed through a variety of estimates to try and get a grip on the number of molecules of different kinds that are in an *E. coli* cell. Why should we care about these numbers? First, a realistic physical picture of any biological phenomenon demands a precise, quantitative understanding of the individual particles involved (for biological phenomena, this usually means molecules) and the spatial dimensions over which they have the freedom to act. One of the most immediate outcomes of our cellular census will be the realization of just how crowded the cellular interior really is, a subject explored in detail in chap. 14. Our census will paint a very different picture of the cellular interior as the seat of biochemical reactions than is suggested by the dilute and homogeneous environment of the biochemical test tube. Indeed, we will see that the mean spacing between protein molecules within a typical cell is less than 10 nm.

Taking the molecular census is also important because we will use our molecular counts in chap. 3 to estimate the rates of macromolecular synthesis during the cell cycle. How fast is a genome replicated? What is the average rate of protein synthesis during the cell cycle and given what we know about ribosomes, how do they maintain this rate of synthesis? A prerequisite to beginning to answer these questions is the macromolecular census itself.

Ultimately, to understand many experiments in biology, it is important to realize that most experimentation is comparative. That is, we compare "normal" behavior to perturbed behavior to see if some measurable property has increased or decreased. To make these statements meaningful, we need to first understand the quantitative baseline relative to which such increases and decreases are compared. There is another sense in which numbers of molecules are particularly meaningful which will be explored in detail in subsequent chapters that has to do with whether we can describe a cell as having "a lot" or "a few" copies of some specific molecule. If a cell has a lot of some particular molecule, then it is appropriate to describe the concentration of that molecule as the basis for predicting cellular function. However, when a cell has only a few copies of a particular molecule, then we need to consider the influence of random chance (or stochasticity) on its function. In many cases, cells have an interesting medium number of molecules where it is not immediately clear which perspective is appropriate. However, knowing the absolute numbers always gives us a reality check for subsequent assumptions and approximations for modeling biological processes.

Because of these considerations, in recent years much effort among biological scientists has been focused on the development of quantitative techniques for measuring the molecular census of living cells (both bacteria and eukaryotes). In this chapter we will rely primarily on order-of-magnitude estimates based on simple assumptions. These estimates are validated by comparison with measurements. In subsequent chapters, these estimates will be refined through explicit model building and direct comparison to quantitative experiments.

- **Estimate: Sizing Up** *E. coli***.** As already noted in the previous chapter, cells are made up of an array of different macromolecules as well as small molecules and ions. To estimate the number of proteins in an *E. coli* cell we begin by noting that with its 1 fL volume, the mass of such a cell is roughly 1 pg, where we have assumed that the density of the cell is that of water which is 1 g/mL. Measurements reveal that the dry weight of the cell is roughly 30 percent of its total and half of that mass is protein. As a result, the total protein mass within the cell is roughly 0.15 pg. We can also estimate the number of carbon atoms in a bacterium on the grounds that roughly half the dry mass comes from the carbon content of these cells, a figure that implies of order 10^{10} carbon atoms per cell. Two of the key sources that have served as a jumping off point for these estimates are Pedersen *et al.* (1978) and Zimmerman and Trach (1991), who describe the result of a molecular census of a bacterium.

  As a first step towards revealing the extent of crowding within a bacterium, we can estimate the number of proteins by assuming a mean protein of 300 amino acids with each amino acid having a characteristic mass of 100 Da. These assumptions are further examined in the problems at the end of the chapter. Using these rules of thumb, we find that the mean protein has a mass of 30,000 Da. Using the conversion factor that 1 Da $\approx 1.6 \times 10^{-24}$ g, we have that our typical protein has a mass of 5×10^{-20} g. The number of proteins per *E. coli* cell is estimated as

$$N_{protein} = \frac{\text{total protein mass}}{\text{mass per protein}} \approx \frac{15 \times 10^{-14} \text{ g}}{5 \times 10^{-20} \text{ g}} \approx 3 \times 10^6. \qquad (2.1)$$

  If we invoke the rough estimate that one-third of the proteins coded for in a typical genome correspond to membrane proteins this implies that the number of cytoplasmic proteins is of order 2×10^6 and the number of membrane proteins is 10^6, although we note that not all of these membrane-associated proteins are strictly transmembrane proteins.

  Another interesting use of this estimate is to get a rough impression of the number of ribosomes - the cellular machines that synthesize proteins. We can estimate the total number of ribosomes by first estimating the total mass of the ribosomes in the cell and then dividing by the mass per ribosome. To be concrete, we need one other fact, which is that roughly 20 percent of the protein complement of a cell is ribosomal protein. If we assume that all of this protein is tied up in assembled ribosomes, then we

can estimate the number of ribosomes by noting: a) that the mass of an individual ribosome is roughly 2.5 MDa and b) that an individual ribosome is roughly 1/3 by mass protein and 2/3 by mass RNA, facts which can be directly confirmed by the reader by inspecting the structural biology of ribosomes. As a result, we have

$$N_{ribosome} = \frac{0.2 \times 0.15 \times 10^{-12} \text{ g}}{830,000 \text{ Da}} \times \frac{1 \text{ Da}}{1.6 \times 10^{-24} \text{ g}} \approx 20,000 \text{ ribosomes.} \tag{2.2}$$

The numerator of the first fraction has 0.2 as the fraction of protein that is ribosomal, 0.15 as the fraction of the total cell mass that is protein and 1 pg as the cell mass. 830,000 Da is our estimate for that part of the ribosomal mass that is protein. The size of a ribosome is roughly 20 nm (in "diameter") and hence the total volume taken up by these 20,000 ribosomes is roughly 10^8 nm^3. This is 10 percent of the total cell volume.

Idealizing an *E. coli* cell as a cube, sphere or spherocylinder yields (see the problems) that the surface area of such cells is $A_{E.coli} \approx 6\ \mu m^2$. This number may be used in turn to estimate the number of lipid molecules associated with the inner and outer membranes of these cells as

$$N_{lipid} \approx \frac{4 \times 0.5 \times A_{E.coli}}{A_{lipid}} \approx \frac{4 \times 0.5 \times (6 \times 10^6 \text{ nm}^2)}{0.5 \text{ nm}^2} \approx 2 \times 10^7, \tag{2.3}$$

where the factor of 4 comes from the fact that the inner and outer membranes are each *bilayers*, implying that the lipids effectively cover the cell surface area four times. A lipid bilayer consists of two sheets of lipids with their tails pointing toward each other. The factor of 0.5 is based on the crude estimate that roughly half of the surface area is covered by membrane proteins rather than lipids themselves. We have made the similarly crude estimate that the area per lipid is 0.5 nm^2. The measured number of lipids is of order 2×10^7 as well.

In terms of sheer numbers, water molecules are by far the majority constituent of the cellular interior. One of the reasons this fact is intriguing is that during the process of cell division, a bacterium such as *E. coli* has to take on a very large number of new water molecules each second. The estimate we do here will be used to examine this transport problem in the next chapter. To estimate the number of water molecules we exploit the fact that roughly 70% of the cellular mass (or volume) is water. As a result, the total mass of water is 0.7 pg. We can find the approximate number of water molecules as

$$N_{H_2O} \approx \frac{0.7 \times 10^{-12} \text{ g}}{18 \text{ g/mole}} \times 6 \times 10^{23} \text{ molecules/mole} \approx 2 \times 10^{10} \text{ water molecules.} \tag{2.4}$$

It is also of interest to gain an impression of the content of inorganic ions in a typical bacterial cell. To that end, we assume that a typical

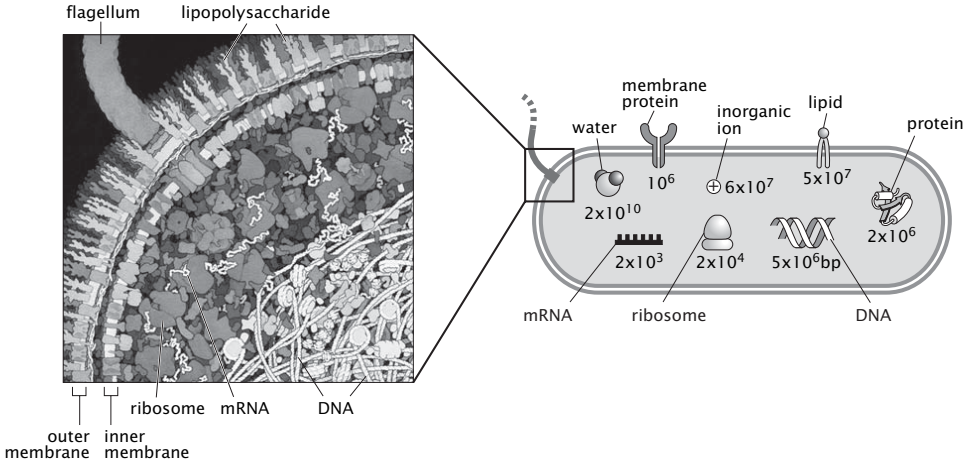

Figure 2.2: Molecular contents of the bacterium *E. coli*. The illustration on the left shows the crowded cytoplasm of the bacterial cell. The cartoon on the right shows an order-of-magnitude molecular census of the *E. coli* bacterium with the approximate number of different molecules in *E. coli*. (Illustration of the cellular interior courtesy of David Goodsell.)

concentration of positively charged ions such as K^+ is 100 mM resulting in the estimate

$$N_{ions} \approx \frac{(100 \times 10^{-3} \text{ moles}) \times (6 \times 10^{23} \text{ molecules/mole})}{10^{15} \ \mu m^3} \times 1 \ \mu m^3 = 6{\times}10^7. \tag{2.5}$$

Here we use the fact that 1 L=15 μm^3. This result could have been obtained even more easily by noting yet another simple rule of thumb, namely, that one molecule per *E. coli* cell corresponds roughly to a concentration of 2 nM.

The outcome of our attempt to size up *E. coli* is illustrated schematically in summary form in fig. 2.2. A more complete census of an *E. coli* bacterium can be found in Neidhardt *et al.* (1990). The outcome of experimental investigations of the molecular census of an *E. coli* cell is summarized (for the purposes of comparing to our estimates) in Table 2.1.2.

How is the census of a cell taken experimentally? This is a question we will return to a number of different times, but will give a first answer here. For the case of *E. coli*, one important tool has been the use of gels like that shown in fig. 2.3. Such experiments work by breaking open cells and keeping only their protein components. The complex protein mixture is then separated into individual molecular species using a polyacrylamide gel matrix. First, the protein mixture is loaded at one end of the gel and an electric field is applied

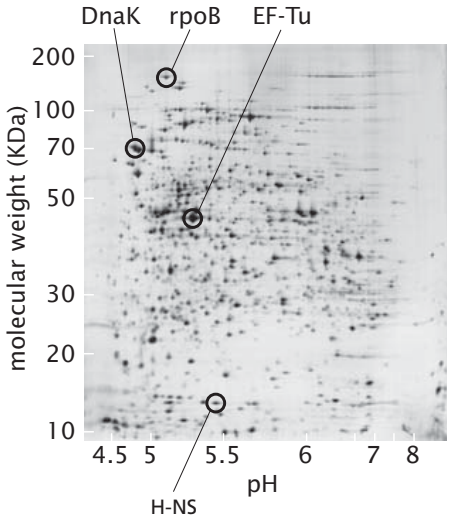

Figure 2.3: Protein census of the cell. Measurement of protein census of *E. coli* using two-dimensional polyacrylamide gel electrophoresis. Each spot represents an individual protein with a unique size and charge distribution. The spots arising from several well known bacterial proteins are labeled. (Adapted from *http://www.expasy.ch/swiss-2dpage/*.)

across the gel, causing the different proteins to migrate through the gel at rates proportional to their net charge. Next, a charged detergent is added that binds to all proteins so the total number of detergent molecules associated with an individual protein is roughly proportional to the protein's overall size, and an electric field is applied in a direction perpendicular to the first one. Because the net charge on the detergent molecules is much larger than the original net charge of the protein, the rate of migration in the second direction through the gel is determined by the protein size. After this procedure, the individual protein species in the original mixture have been resolved into a series of spots on the gel, with large, negatively charged proteins at the upper lefthand corner and small, positively charged proteins at the lower righthand corner for the gel shown in fig. 2.3. The proteins can then be stained with a nonspecific dye so that there locations within the gel can be directly observed. The intensity of the spots on such a gel can then be used as a basis for quantifying each species. The identity of the protein that congregates in each spot can then be determined by physically cutting each spot out of the gel, eluting the protein and determining its size and amino acid content using mass spectroscopy. Similar tricks are used to characterize the amount of RNA and lipids, for example, resulting in a total census like that shown in Table 2.1.2.

The Cellular Interior Is Highly Crowded With Mean Spacings Between Molecules That Are Comparable to Molecular Dimensions

Substance	% of total dry weight	Number of molecules
Macromolecule		
Protein	55.0	2.4×10^6
RNA	20.4	
23S RNA	10.6	19,000
16S RNA	5.5	19,000
5S RNA	0.4	19,000
Transfer RNA (4S)	2.9	200,000
Messenger RNA	0.8	1,400
Phospholipid	9.1	22×10^6
Lipopolysaccharide	3.4	1.2×10^6
DNA	3.1	2
Murein	2.5	1
Glycogen	2.5	4,360
Total macromolecules	**96.1**	
Small molecules		
Metabolites, building blocks, etc.	2.9	
Inorganic ions	1.0	
Total small molecules	**3.9**	

Table 2.1: Observed macromolecular census of an *E. coli* cell. (Data from F. C. Neidhardt *et al.*, Physiology of the Bacterial Cell, Sunderland, Sinauer Associates Inc., 1990 and M. Schaechter *et al.*, Microbe, Washington DC, ASM Press, Washington, 2006.

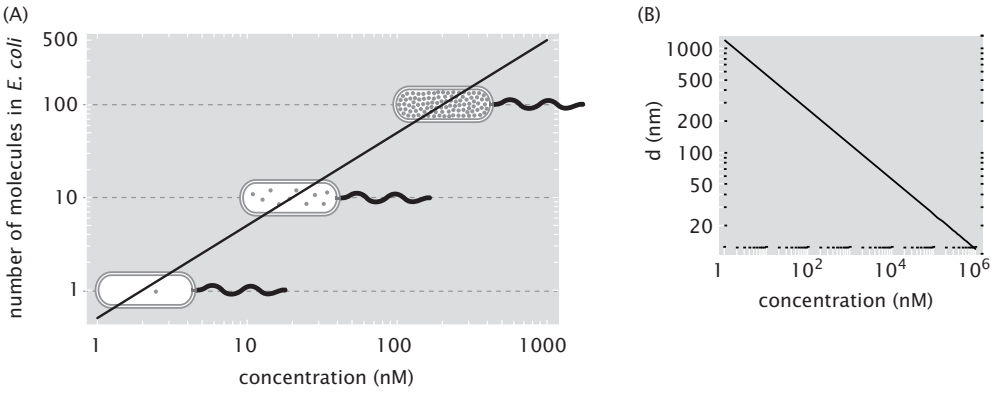

Figure 2.4: Physical interpretation of concentration. (A) Concentration in *E. coli* units. Number of copies of a given molecule in a volume the size of an *E. coli* cell as a function of the concentration. (B) Concentration expressed in units of typical distance (d) between neighboring molecules measured in nanometers.

One of the most intriguing implications of our census of the molecular parts list of a bacterium is the extent to which the cellular interior is crowded. Because of experiments and associated estimates on the contents of *E. coli*, it is now possible to construct illustrations to depict the cellular interior in a way that respects the molecular census. The crowded environs of the interior of such a cell is shown in fig. 2.2. This figure gives a number of different views of the crowding associated with any *in vivo* process. In chap. 14, we will see that this crowding effect will force us to call in question our simplest models of chemical potentials, the properties of water and the nature of diffusion. The generic conclusion is that the mean spacing of proteins and their assemblies is comparable to the dimensions of these macromolecules themselves. The cell is a very crowded place!

The quantitative significance of fig. 2.2 can be further appreciated by converting these numbers into concentrations. To do so, we recall that the volume of an *E. coli* cell is 1 fL. The rule of thumb that emerges from this analysis is that 2 nM implies roughly one molecule per bacterium. A concentration of 2 μM implies roughly 1000 copies of that molecule per cell. Concentration in terms of our standard ruler is shown in fig. 2.4. This figure shows the number of copies of the molecule of interest in such a cell as a function of the concentration.

We can use these concentrations directly to compute the mean spacing between molecules. That is, given a certain concentration, there is a corresponding average distance between the molecules. Having a sense of this distance can serve as a guide to thinking about the likelihood of diffusive encounters and reactions between various molecular constituents. If we imagine the molecules at a given concentration arranged on a cubic lattice of points, then the mean

spacing between those points is given by

$$d = c^{-1/3}, \tag{2.6}$$

where c is the concentration of interest (measured in units of number of molecules per unit volume). Larger concentrations imply smaller intermolecular spacings. This idea is formalized in fig. 2.4 which shows the relation between the mean spacing measured in nanometers and the concentration.

2.1.3 Looking Inside of Cells

With our reference bacterium in mind, the remainder of the chapter focuses on the various structures that make up cells and organisms. To talk about these structures, it is helpful to have a sense of how we know what we know about them. Further, model building requires facts. To that end, we periodically take stock of the experimental basis for our models. For this chapter, the "Experiments Behind the Facts" focuses on how we know what we know about biological structures.

- **Experiments Behind the Facts: Probing Biological Structure.** To size up cells and their organelles we need to extract "typical" structural parameters from a variety of experimental studies. Though we leave a description of the design and setup of such experiments to more specialized texts, the goal is to provide at least enough details that the reader sees where some of the key structural facts that we will use throughout the book come from. We emphasize two broad categories of experiments: i) those in which some form of radiation interacts with the structure of interest and ii) those in which forces are applied to the structure of interest.

  Fig. 2.5 shows three distinct experimental strategies which feed into our estimates and all of which reveal different facets of biological structure. One of the mainstays of structural analysis is light microscopy. Fig. 2.5(A) shows a schematic of the way in which light can excite fluorescence in samples that have some distribution of fluorescent molecules within them. In particular, this example shows a schematic of a microtubule which has some distribution of fluorophores along its length. Incident photons of one wavelength are absorbed by the fluorophore and this excitation leads them to emit light of a different wavelength which is then detected. As a result of selective labeling of only the microtubules with fluorophores, when examined in the microscope it is only these structures that are observed. These experiments permit a determination of the size of various structures of interest, how many of them there are and where they are localized. By calibrating the intensity from single fluorophores it has become possible to take a single molecule census for many of the important proteins in cells. For an example of this strategy, see Wu and Pollard, 2005.

  A totally different window on the structure of the cell and its components is provided by tools such as the atomic-force microscope (AFM). As will

be explained in chap. 10, the AFM is a cantilever beam with a sharp tip on its end. The tip is brought very close to the surface where the structure of interest is present and is then scanned in the plane. One way to operate the instrument is to move the cantilever up and down so that the force applied on the tip remains constant. Effectively, this demands a continual adjustment of the height as a function of the x-y position of the tip. The nonuniform pattern of cantilever displacements can be used to map out the topography of the structure of interest. Fig. 2.5(B) shows a schematic of an atomic-force microscope scanning a typical fibroblast cell as well as a corresponding image of the cell.

Fig. 2.5(C) gives a schematic of the way in which x-rays or electrons are scattered off of a biological sample. The schematic shows an incident plane wave of radiation which interacts with the biological specimen and results in the emergence of radiation with the same wavelength but a new propagation direction. Each point within the sample can be thought of as a source of radiation and the observed intensity at the detector reflects the interference from all of these different sources. By observing the pattern of intensity it is possible to deduce something about the structure that did the scattering. This same basic idea is applicable to a wide variety of radiation sources including x-rays, neutrons and electrons.

An important variation on the theme of measuring the scattered intensity from irradiated samples is cryo-electron tomography. This technique is one of the centerpieces of structural biology and is built around uniting electron microscopy with sample preparation techniques which rapidly freeze the sample. The use of tomographic methods has made it possible to go beyond the planar sections seen in conventional electron microscopy images. The basic idea of the technique is indicated schematically in fig. 2.6, and is built around the idea of rotating the sample over a wide range of orientations and then to build up a corresponding three-dimensional reconstruction on the basis of the entirety of these images. These techniques have already revolutionized our understanding of particular organelles and are now being used to image entire cells.

2.1.4 Where Does *E. coli* Fit?

Biological Structures Exist Over a Huge Range of Scales

The spatial scales associated with biological structures run from the nanometer scale of individual molecules, all the way to the scale of the Earth itself. Where does *E. coli* fit into this hierachy of structures? Fig. 2.7 shows the different structures that can be seen as we scale in and out from an *E. coli* cell. At each scale, new classes of structure can be seen. A roughly tenfold increase in magnification relative to an individual bacterium reveals the viruses that attack bacteria. These viruses, known as bacteriophage, have a characteristic scale of

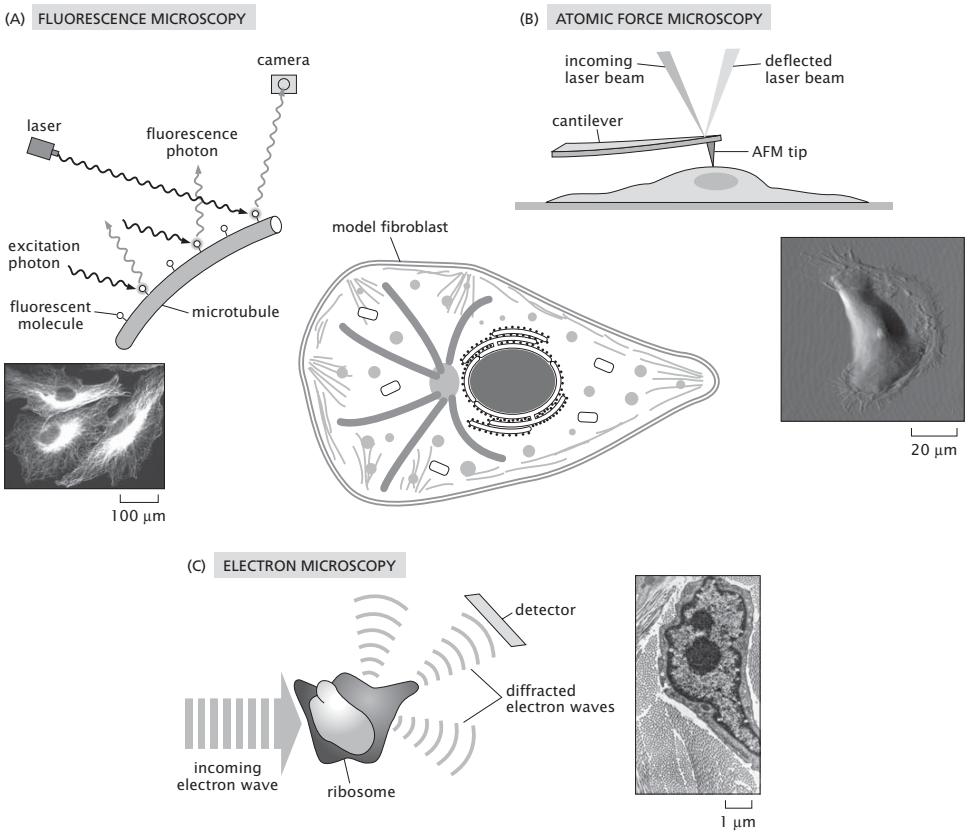

Figure 2.5: Experimental techniques which have revealed the structure of both cells and their organelles. (A) Fluorescence microscopy and associated image of fibroblast with labeled microtubules, (B) Atomic force microscopy schematic and associated image of surface topography of fibroblast. (C) Electron microscopy schematic and images of a fibroblast. (A, adapted from W. Yu *et al.* J. Neurosci. 25:5573, 2005; B, courtesy of M. Radmacher; C, adapted from P. C. Cross and K. L. Mercer, Cell and Tissue Ultrastructure, New York: W. H. Freeman and Company, 1993.)

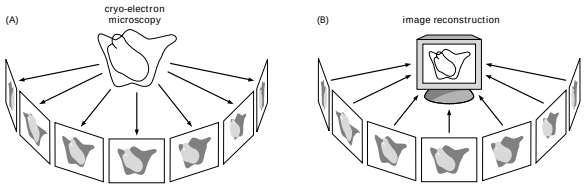

Figure 2.6: Schematic of tomographic reconstruction. (A) The sample is rotated so that radiation is scattered from a series of different orientations, (B) three-dimensional reconstruction of the structure giving rise to the pattern of scattering.

roughly 100 nm. They are made up of a protein shell (the capsid) which is filled with the viral genome. Continuing our downward descent using yet higher magnification, we see the ordered packing of the viral genome within its capsid. These structures are intriguing because they involve the ordered arrangement of more than 10 μm of DNA in a capsid which is less than 100 nm across. Another rough factor of ten increase in resolution reveals the structure of the DNA molecule itself with a characteristic cross sectional radius of roughly 1 nm and a length of 3.4 nm per helical repeat.

A similar scaling out strategy reveals new classes of structures. As shown in fig. 2.7, a tenfold increase in spatial scale brings us to the realm of eukaryotic cells in general, and specifically, to the scale of the epithelial cells that line the human intestine. We use this example because bacteria such as *E. coli* are a central player as part of our intestinal fauna. Another tenfold increase in spatial scale reveals one of the most important inventions of evolution, namely, multicellularity. In this case, the cartoon depicts the formation of planar sheets of epithelial cells. These planar sheets are themselves the building blocks of yet higher-order structures such as tissues. Scaling out to larger scales would bring us to multicellular organisms and the structures they build.

The remainder of the chapter is devoted to an attempt to take stock of the structures at each of these scales and to provide a feeling for the molecular building blocks that make up these different structures. Our strategy will be to build upon our cell-centered view and to first descend in length scale from that of cells to the molecules they are made of. Once this structural descent is complete, we will embark on an analysis of biological structure in which we zoom out from the scale of individual cells to collections of cells.

2.2 Cells and Structures Within Them

2.2.1 Cells: A Rogue's Gallery

All living organisms are based on cells as the indivisible unit of biological organization. However, within this general rule there is tremendous diversity among

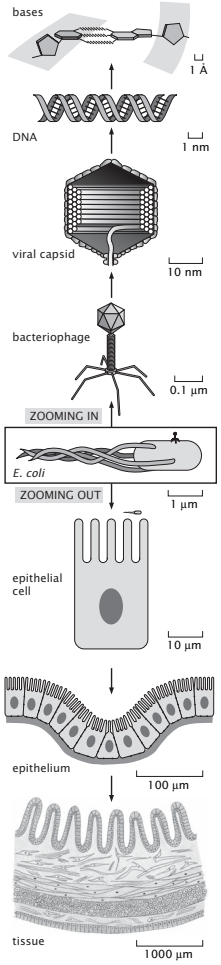

Figure 2.7: Powers of ten representation of biological length scales. The hierarchy of scales is built around the *E. coli* standard ruler. Starting with *E. coli* the first part of the chapter will consider a succession of tenfold increases in resolution as are shown in the figure. The second part of the chapter will zoom out from the scale of an *E. coli* cell.

living cells. Several billion years ago, our last common ancestor gave rise to three different lineages of cells now commonly called Bacteria, Archaea and Eukarya, a classification suggested by similarities and differences in ribosomal RNA sequences. Every living organism on Earth is a member of one of these groups. Most bacteria and archaea are small (3 μm or less) and extremely diverse in their preferred habitats and associated lifestyles ranging from geothermal vents at the bottom of the ocean to permafrost in Antarctica. Bacteria and archaea look similar to one another and it has only been within the last few decades that molecular analysis has revealed that they are completely distinct lineages that are no more closely related to each other than the two are to eukaryotes.

The organisms that we encounter in our everyday life and can see with the naked eye are members of Eukarya (individuals are called eukaryotes). These include all animals, all plants ranging from trees to moss and also all fungi such as mushrooms and mold. Thus far we have focused on *E. coli* as a representative cell although we must acknowledge that *E. coli*, as a member of the bacterial group, is in some ways very different from a eukaryotic or archaeal cell. The traditional definition of a eukaryotic cell is one that contains its DNA genome within a membrane-bound nucleus. Most bacteria and archaea lack this feature and also lack other elaborate intracellular membrane-bound structures such as the endoplasmic reticulum and the Golgi apparatus that are characteristic of the larger and more complex eukaryotic cells.

Cells Come in a Wide Variety of Shapes and Sizes and With a Huge Range of Functions

Cells come in such a wide variety of shapes, sizes and lifestyles that choosing one representative cell type to tell their structural story is misleading. In fig. 2.8 we show a rogue's gallery illustrating the variety of cell sizes and shapes found in the eukaryotic group, all referenced to the *E. coli* standard ruler. This gallery is by no means complete. There is much more variety than we can illustrate, but this covers a reasonable range of eukaryotic cell types that have been well studied by biologists. In this figure we have chosen a variety of examples that represent experimental bias among biologists where more than half of the examples are human cells and the others represent the rest of the eukaryotic group. The vast majority of eukaryotes are members of a group called protists. This poorly-defined group encompasses all eukaryotes that are neither plants nor animals nor fungi. Protists are extremely diverse in their appearance and lifestyles, but they are all small (ranging from 0.002 mm to 2 mm). Some examples of protists include marine diatoms such as *Emiliana huxleyi*, soil amoeba such as *Dictyostelium discoideum* and the lovely creature *Paramecium* seen in any sample of pond water and familiar from many high-school biology classes. Another notable protist is the pathogen that causes malaria called *Plasmodium falciparum*. Fig. 2.8(A) shows the intriguing protist *Giardia lamblia*, a parasite known to hikers as a source of water contamination. A broader set of examples is shown in fig. 20.3 (pg. 1086).

Although protists constitute the vast majority of eukaryotic cells on the planet, biologists are often inclined to study cells more related to us. This

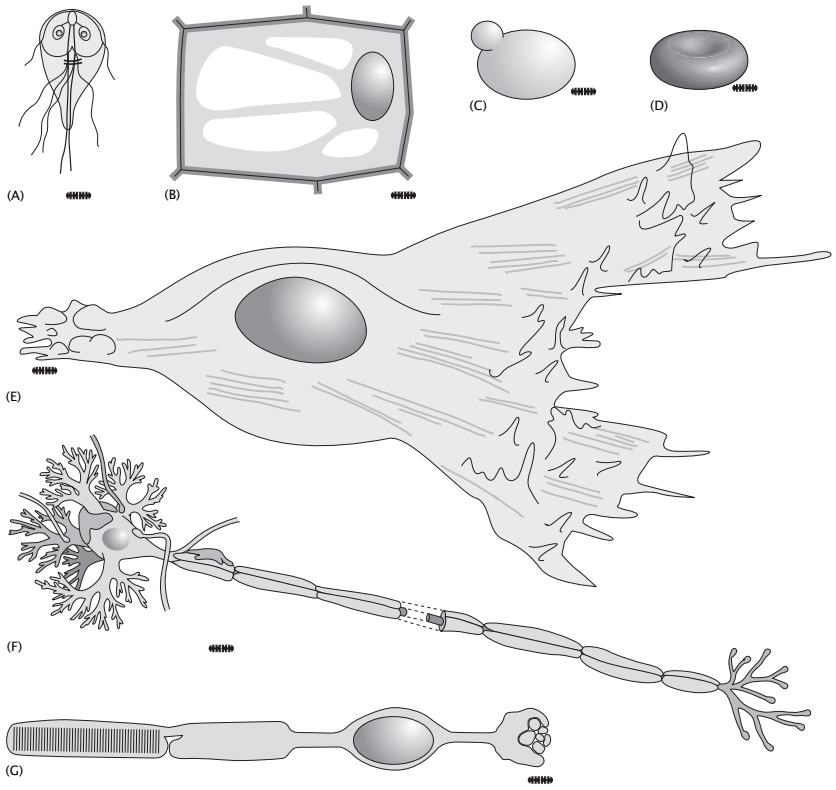

Figure 2.8: Cartoons of several different types of cells all referenced to the standard *E. coli* ruler. (A) the protist *Giardia lamblia*, (B) plant cell, (C) budding yeast cell, (D) red blood cell, (E) fibroblast cell, (F) eukaryotic nerve cell and (G) rod cell.

includes the plant kingdom which is obviously important as a source of food and flowers. Plant cells like that shown in fig. 2.8(B) are characterized by a rigid cell wall, often giving them angular structures like that shown in the figure. The typical length scale associated with these cells is often tens of microns. One of the distinctive features of these cell is their large vacuoles within the intracellular space that hold water and contribute to the mechanical properties of plant stems. These large vacuoles are very distinct from animal cells where most of the intracellular space is filled with cytoplasm. Consequently, in comparing a plant and animal cell of similar overall size, the plant cell will have typically tenfold less cytoplasmic volume because most of its intracellular space is filled with vacuoles. Hydrostatic forces matter much more to plants than animals. For example, a wilting flower can be revived simply by application of water since this allows the vacuoles to fill and stiffen the plant stem.

Among all of the eukaryotes, the group most closely related to the animals (as evidenced by ribosomal RNA similarity, for example) is, surprisingly, the fungi. The representative fungus shown in the figure is the budding yeast *Saccharomyces cerevisiae* (which we will refer to as *S. cerevisiae*). *S. cerevisiae* was domesticated by humans several thousand years ago and continues to serve as a treasured microbial friend that makes our bread rise and provides alcohol in our fermented beverages such as wine. Just as *E. coli* sometimes serves as a key model prokaryotic system, the yeast cell often serves as the model single-celled eukaryotic organism. Besides the fact that humans are fond of *S. cerevisiae* for its own intrinsic properties, it is also useful to biologists as a representative fungus. Of all the other organisms on Earth, fungi are closest to animals in terms of evolutionary descent and similarity of protein functions. Although there are no single-celled animals, there are some single celled fungi including *S. cervisiae*. Therefore, many laboratory biological experiments relying on rapid replication of single cells are most easily performed on this organism and its relatives *Candida albicans* and *Schizosaccharomyces pombe* (a "fission" yeast that divides in the middle). Fig. 2.9 shows a scanning electron microscope image of a yeast cell engaged in budding. As this image shows, the geometry of yeast is relatively simple compared to many other eukaryotic cells and it is also a fairly small member of this group with a characteristic diameter of roughly 5 microns. Nonetheless, it possesses all the important structural hallmarks of the eukaryotes including, in particular, a membrane bound nucleus, segregating the DNA genome from the cytoplasmic machinery that performs most metabolic function.

Earlier, we estimated the molecular census of an *E. coli* cell. It will now be informative to compare those estimates with the corresponding model eukaryotic cell that will continue to serve as a comparative basis for all our eukaryotic estimates.

- **Estimate: Sizing Up Yeast.** The volume of a yeast cell can be computed in *E. coli* volume units, $V_{E.\ coli}$. In particular, if we recall that $V_{E.\ coli} \approx 1.0\ \mu\text{m}^3$ and think of yeast as a sphere of diameter 5 μm, then we have the relation $V_{yeast} \approx 60 V_{E.\ coli}$, that is, roughly 60 *E. coli* cells

would fit inside of a yeast cell. The surface area of a yeast cell can be estimated using a radius of $r_{yeast} \approx 2.5\ \mu m$ which yields $A_{yeast} \approx 80\ \mu m^2$. If we treat the yeast nucleus as a sphere with a diameter of roughly 2.0 μm, its volume is roughly 4 μm^3. Within this nucleus is housed the 1.2×10^7 bp of the yeast genome which is divided amongst 16 chromosomes. The DNA in yeast is packed into higher order structures mediated by protein assemblies known as histones. In particular, the DNA is wrapped around a series of cylindrical cores made up of eight such histone proteins each, with roughly 150 bp wrapped around each histone octamer, and approximately a 50 bp spacer between. As a result, we can estimate the number of nucleosomes (the histone-DNA complex) as

$$N_{nucleosome} \approx \frac{12 \times 10^6 \text{ bp}}{200 \text{ bp / nucleosome}} \approx 60,000. \tag{2.7}$$

Experimentally, the measured number appears to be closer to 80,000, with a mean spacing between nucleosomes of order 170 bp. The total volume taken up by the histones is roughly 230 nm^3 per histone (thinking of each histone octamer as a cylindrical disk of radius 3.5 nm and height 6 nm), resulting in a total volume of 14×10^6 nm^3 taken up by the histones. The volume taken up by the genome itself is comparable at 1.2×10^7 nm^3, where we have used the rule of thumb that the volume per base pair is 1 nm^3. The packing fraction (defined as the ratio of the volume taken up by the genome to the volume of the cell) associated with the yeast genomic DNA can be estimated by evaluating the ratio

$$\rho_{pack} \approx \frac{(1.2 \times 10^7 \text{ bp}) \times (1 \text{ nm}^3/\text{bp})}{4 \times 10^9 \text{ nm}^3} \approx 3 \times 10^{-3}. \tag{2.8}$$

Note that we have used the fact that the yeast genome is 1.2×10^7 base pairs in length and is packed in the nucleus which has a volume of ≈ 4 μm^3.

These geometric estimates may be used to make corresponding molecular estimates such as the number of lipids and proteins in a typical yeast cell. The number of proteins can be estimated several ways - perhaps the simplest is just to assume that the fractional occupancy of yeast cytoplasm is identical to that of *E. coli* with the result that there will be 60 times as many proteins in yeast as in *E. coli* based strictly on scaling up the cytoplasmic volume. This simple estimate is obtained by *assuming* that the composition of the yeast interior is more or less the same as that of an *E. coli* cell. This strategy results in

$$N_{protein}^{yeast} \approx 60 \times N_{protein}^{E.coli} \approx 2 \times 10^8. \tag{2.9}$$

The number of lipids associated with the plasma membrane of the yeast cell can be obtained as

$$N_{lipid} \approx \frac{2 \times 0.5 \times A_{yeast}}{A_{lipid}} \approx \frac{2 \times 0.5 \times (80 \times 10^6 \text{ nm}^2)}{0.5 \text{ nm}^2} \approx 2 \times 10^8, \tag{2.10}$$

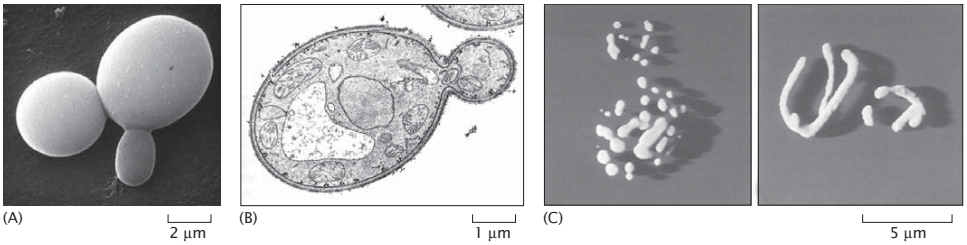

Figure 2.9: Microscopy images of yeast and their organelles. (A) Scanning electron micrograph of the yeast *Candida albicans* revealing the overall size scale of budding yeast. (B) Electron microscopy image of a budding *Candida albicans* yeast cell. (C) and (D) Confocal microscopy images of the mitochondria of *Saccharomyces cerevisiae*. (A, adapted from *http://overcomingcandida.com/candida_albicans_pictures.htm*; B, adapted from G. M. Walker, Yeast, Physiology and Biotechnology, Chichester, John Wiley and Sons, 1998; C, D, adapted from W. Visser *et al.*, Antonie van Leeuwenhoek 67:243, 1995.)

where the factor of 0.5 is based on the idea that roughly half of the surface area is covered by membrane proteins rather than lipids themselves and the factor of 2 accounts for the fact that the membrane is a bilayer. This should be contrasted with the situation in *E. coli* which has a double-membrane surrounding the cytoplasm.

Another interesting estimate suggested by fig. 2.9(C) is associated with the organellar content of these cells. In particular, this figure shows the mitochondria of yeast which are being grown in two different media. These pictures suggest several interesting questions such as what fraction of the cellular volume is occupied by mitochondria and what is the surface area tied up with the mitochondrial outer membranes? The number of mitochondria in the image can be estimated several ways - one of which is to attempt to count them directly, the other of which is to estimate their mean spacing and to compute the corresponding density and number. Using the latter method results in an estimate of roughly 40 mitochondria in the image of fig. 2.9(C). Further, we estimate that the typical mitochondrial size is roughly 3/4 μm, resulting in a total mitochondrial volume of

$$V_{mito} \approx 40 \times \frac{4\pi}{3}(\frac{3}{8})^3 \ \mu m^3 \approx 9 \ \mu m^3, \tag{2.11}$$

which given the total volume of the cell of 60 μm^3 translates into a volume fraction of roughly 15 percent. The total area of the outer membranes of these mitochondria is roughly 70 μm^2, comparable to the entire area of the plasma membrane itself. The analysis of the image in fig. 2.9(D) is left as an exercise for the reader in the problems.

Our estimates are brought into sharpest focus when they are juxtaposed with actual measurements. The census of yeast cells has been performed in several distinct and fascinating ways in recent years. The key idea is to generate thousands of different yeast strains, each of which has a tag on a different one of the yeast gene products. For example, it is possible to generate strains with a peptide fragment that can then be recognized by antibodies. A second scheme is to construct protein fusions in which the protein of interest is attached to a fluorescent protein such as the green fluorescent protein (GFP). Then, by querying each and every cell either by examining the extent of antibody binding or fluorescence, it is possible to count up the numbers of each type of protein. Fig. 2.10(B) shows a histogram of the number of proteins that occur with a given protein copy number. Although the average protein appears to be present in the cell with a few thousand copies, this histogram shows that some proteins are present with fewer than fifty copies and others with more than a million copies. Further, similarly quantifying mRNA as shown in fig. 2.10(C) reveals that many genes including essential genes are expressed with an average of less than one molecule of RNA per cell. By adding up the total number of proteins on the basis of this census, we estimate there are 50×10^6 proteins in a yeast cell, somewhat less than suggested by our crude estimate given above. Now that we have completed our first introduction to the important yeast as a representative eukaryote, we return to our tour of cell types in fig. 2.8.

Cells From Humans Have a Huge Diversity of Structure and Function

The remainder of the cells in fig. 2.8 are all human cells and show another interesting aspect of cellular diversity. To a first approximation, every cell in the human body contains the same DNA genome. And yet, individual human cells differ significantly with respect to their sizes (with sizes varying from roughly 5 microns to 1 meter for the largest neurons), shapes and functions. For example, rod cells in the retina are specialized to detect incoming light and transmit that information to the neural system so that we can see. Red blood cells are primarily specialized as carriers of oxygen and, in fact, are dramatically different from almost all other cells in having dispensed with their nucleus as part of their developmental process. As we will discuss extensively throughout the book, other cells have other unique classes of specialization.

Fig. 2.8(D) is a schematic of the structure of a red blood cell. Note that the shapes of these cells are decidedly not spherical raising interesting questions about the mechanisms of cell-shape maintenance. Despite their characteristic size of order 5 microns, these cells easily pass through capillaries with less than half their diameter as shown in fig. 2.11, implying that their shape is altered significantly as part of their normal life cycle. While in capillaries (either artificial or *in vivo*), the red blood cell is severely deformed to pass through the narrow passage. In their role as the transport vessels for oxygen-rich hemoglobin, these cells will serve as an inspiration for our discussion of the statistical mechanics of cooperative binding. Red blood cells are a target of one of the most common infectious diseases suffered by humans caused by the invasion of a protozoan. Malaria infected red blood cells are much stiffer than normal cells and cannot

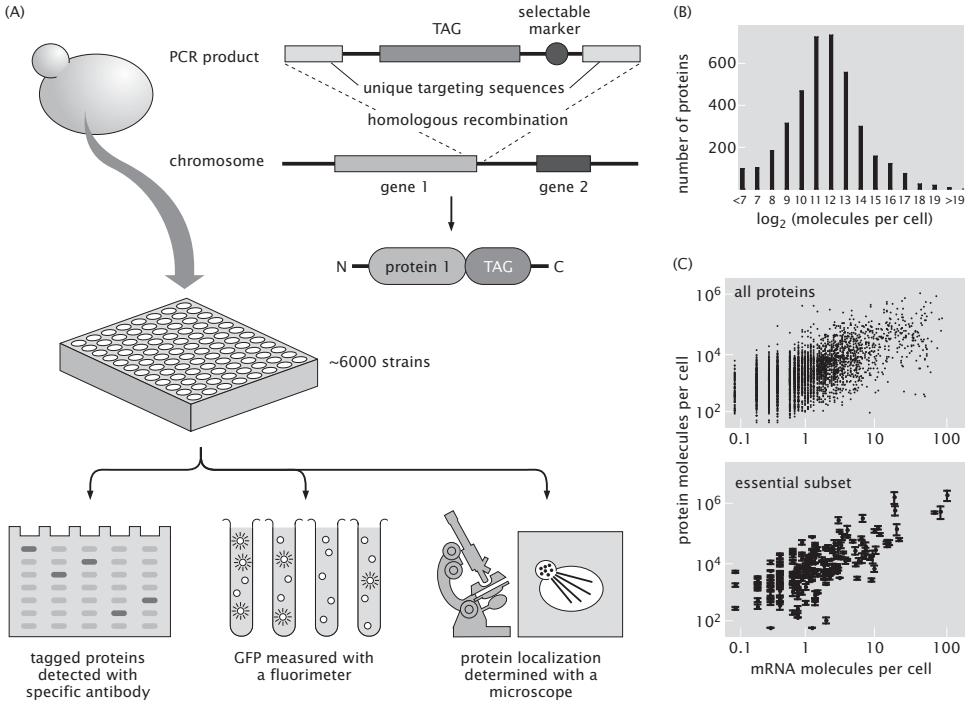

Figure 2.10: Protein copy numbers in yeast. (A) Schematic of constructs used to measure the protein census and several different methods for quantifying protein in labeled cells. (B) Result of antibody detection of various proteins in yeast showing the number of proteins that have a given copy number. The number of copies of the protein is expressed in powers of 2 as 2^N. (C) Abundances of various proteins as a function of their associated mRNA copy numbers. The bottom plot shows this result for essential soluble proteins. (A-C, adapted from S. Ghaemmaghami *et al.* Nature, 425:737, 2003.)

deform to enter small capillaries. Consequently, people suffering from malaria experience severe pain and damage to tissues because of the inability of their red blood cells to enter those tissues and deliver oxygen.

One of the favorite eukaryotic cells from multicellular organisms is the fibroblast as shown schematically in fig. 2.8(E) and shown in an AFM image in fig. 2.12. These cells will serve as a centerpiece for much of what we will have to say about "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 (fig. 2.8(F)) and rod cells (fig. 2.8(G)) reveal a great deal more complexity than the examples highlighted above. In these cases, the cell shape is intimately related to their 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 which 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 tens of centimeters or even more. 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. Fig. 2.8 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. Fig. 2.13 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, serving 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 the fibroblast as an example, then the cell itself has dimensions of roughly 50 microns, while the nucleus has a character-

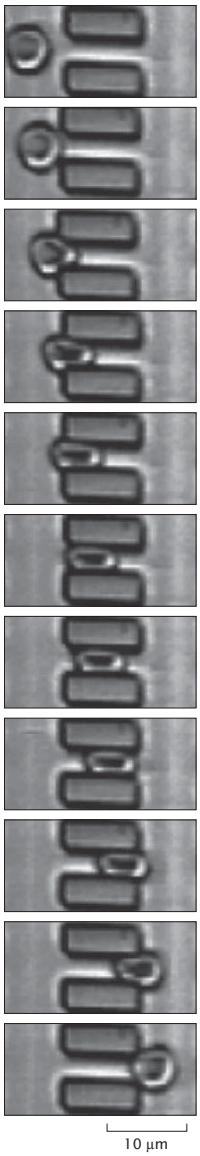

Figure 2.11: Deformability of red blood cells. To measure the deformability of human red blood cells, an array of blocks was fabricated in silicon, each block was $4 \times 4 \times 12$ microns. The blocks were spaced by 4 microns in one direction and 13 microns in the other. A glass coverslip covered 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 throught 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 James Brody.)

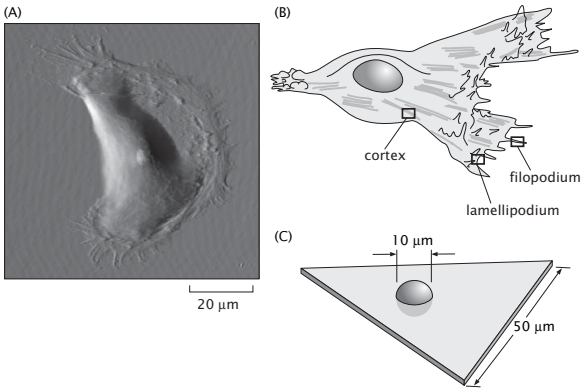

Figure 2.12: Structure of a fibroblast. (A) Atomic-force microscopy image of a fibroblast, (B) cartoon of the external morphology of a fibroblast, (C) characteristic dimensions of a "typical" fibroblast. (A, courtesy of Manfred Radmacher.)

istic linear dimension of roughly 10 microns as shown schematically in fig. 2.12. 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 chap. 8. Transcription occurs in the nucleus as well as several kinds of RNA processing. 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 membraneous 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 fig. 2.14, 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. From images such as those in fig. 2.14 and 2.15 it is clear that the ER can assume different geometries in different cell types and under different conditions. How much total membrane area is taken up by the ER? 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 fig. 2.14 which suggests that we think of the ER as a series of concentric spheres centered about the

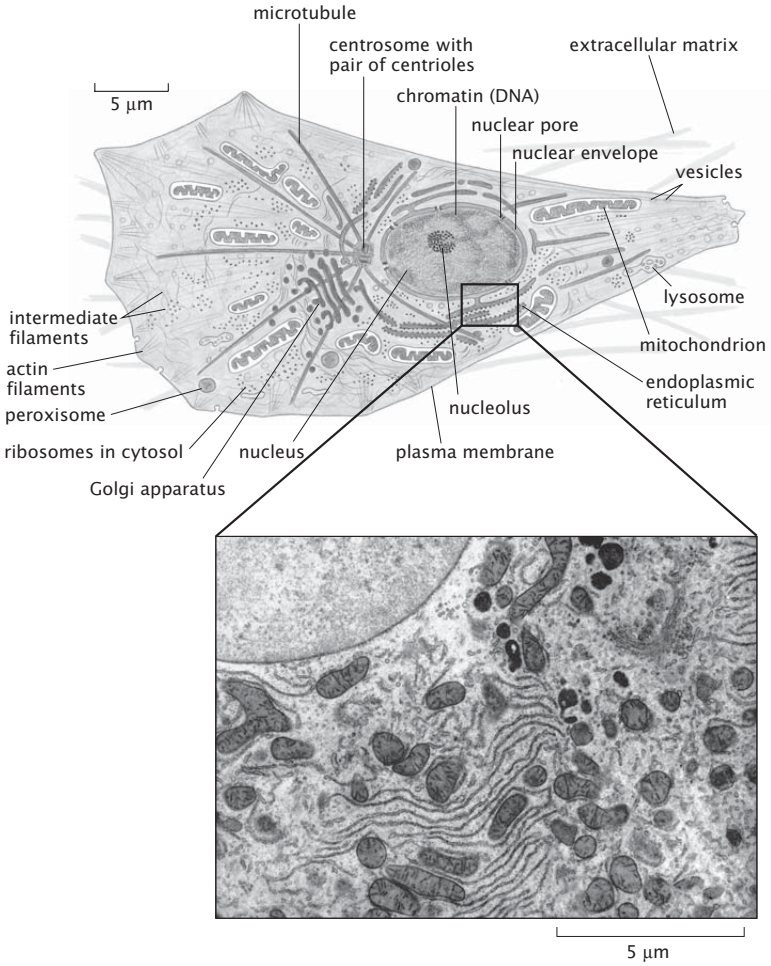

Figure 2.13: Eukaryotic cell and its organelles. The schematic shows a eukaryotic cell and a variety of membrane bound organelles. A thin-section electron microscopy image shows a portion of a rat liver cell approximately equivalent to the boxed area on the schematic. A portion of the nucleus can be seen in the upper left corner. The most prominent organelles visible in the image are mitochondria, lysosomes, the rough endoplasmic reticulum and the Golgi apparatus. (Electron micrograph from D. W. Fawcett, The Cell: An Atlas of Fine Structure, Philadelphia: W. B. Saunders & Co., 1966.)

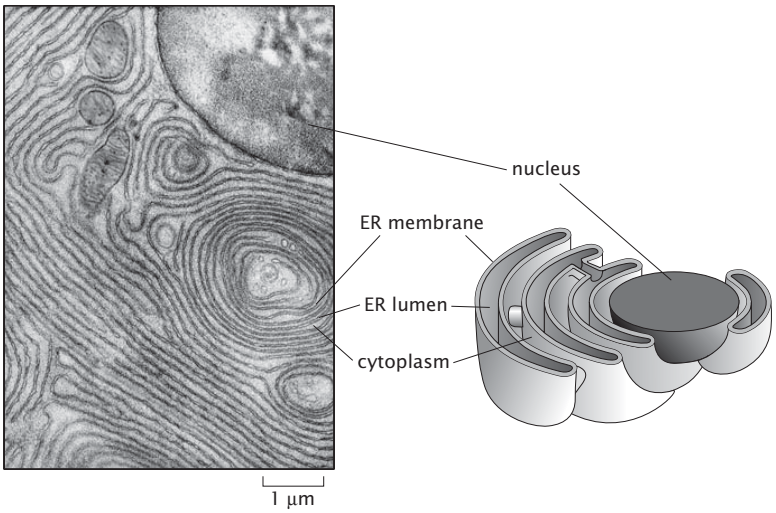

Figure 2.14: Electron micrograph and associated schematic of the endoplasmic reticulum. The left panel shows a thin-section electron micrograph of an acinar cell from the pancreas of a bat. The nucleus is visible at the upper right and the dense and elaborate ER structure is strikingly evident. The right panel shows a schematic diagram of a model for the three-dimensional structure of the ER in this cell. Notice that the size of the lumen in the ER in the schematic is exaggerated for ease of interpretation. (Electron micrograph from D. W. Fawcett, The Cell: An Atlas of Fine Structure, Philadelphia: W. B. Saunders & Co., 1966.)

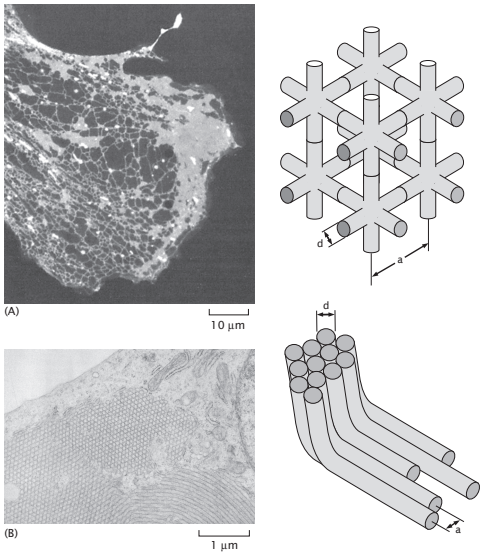

Figure 2.15: Variable morphology of the ER. (A) In most cultured cells, the ER is a combination of a web-like reticular network of tubules and larger flattened cisternae. In this image, a cultured fibroblast was stained with a fluorescent dye called DiOC6 that specifically labels ER membrane. On the right is a schematic of an idealized three-dimensional reticular network. (B) Some specialized cells and those treated with drugs that up regulate the synthesis of lipids reorganize their ER 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. Nat. Acad. Sci., 79:1185, 1982.)

nucleus. We follow Fawcett (1966) who characterizes the ER as forming "lamellar systems of flat cavities, rather uniformly spaced and parallel to one another" as shown in fig. 2.14.

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

$$V_{\rm ER} = \sum_i A_i d \ , \tag{2.12}$$

where A_i is the area of the $i^{\rm th}$ concentric sphere and d is the distance between adjacent cisternae. Since two membranes bound each cisterna the total area of the ER membrane is $A_{\rm ER} = 2 \times \sum_i A_i$. In our model, the total volume of the ER 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

$$V_{\rm ER} = \frac{4\pi}{3} R_{out}^3 - \frac{4\pi}{3} R_{nucleus}^3 \ . \tag{2.13}$$

Combining the two ways of computing the volume of the ER, eqns. 2.12 and 2.13, we arrive at an expression for the ER area,

$$A_{\rm ER} = \frac{8\pi}{3d}(R_{out}^3 - R_{nucleus}^3) \ . \tag{2.14}$$

Using the values $R_{\rm nucleus} = 5$ μm, $R_{\rm out} = 10$ μm and $d = 0.05$ μm, we get at an estimate $A_{\rm ER} = 15 \times 10^4$ $\mu{\rm m}^2$. This result should be contrasted with a crude estimate for the area of a fibroblast which can be obtained by using the dimensions in fig. 2.12(C) and which yields an area of 10^4 $\mu{\rm m}^2$ for the cell membrane itself. To estimate the area of the ER when it is in reticular form we describe its structure as interpenetrating cylinders of diameter $d \approx 10$ nm separated by a distance $a \approx 60$ nm, as shown in fig. 2.15. The completion of the estimate is left to the problems, but results in a comparable membrane area.

The other major organelles found in most cells and visible in fig. 2.13 include the Golgi apparatus, mitochondria and lysosomes. The Golgi apparatus, similar to the ER, is largely involved in processing and trafficking of membrane-bound and secreted proteins (for another view, see fig. 11.34 on pg. 605). 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 fig. 11.35 on pg. 606). The mitochondria are the primary site of ATP synthesis for cells growing in the presence of oxygen, and their physiology as well as their structure are fascinating and have been well studied. We

will return to the topic of mitochondrial structure in chap. 11 and discuss the workings of the tiny machine responsible for ATP synthesis in chap. 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.

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 and as indicated schematically in fig. 2.8 (pg. 72). The common theme is that all organelles represent 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 (Somes)

Proteins, nucleic acids, sugars and lipids often work as a team. Indeed, as will become clear throughout the remainder of the book, these macromolecules often come together to make assemblies, often dubbed "somes". We think of yet another factor of ten magnification relative to the previous section, and with this increase of magnification we see assemblies such as those shown in cartoon form in fig. 2.16. The genetic material in the eukaryotic nucleus is organized into chromatin fibers which themselves are built up of protein-DNA assemblies known as nucleosomes (fig. 2.16(A)). Traffic of molecules such as completed RNA molecules out of the nucleus is mediated by the elaborate nuclear pore complex (fig. 2.16(B)). The replication complex that copies DNA before cell division is similarly a collection of molecules which has been dubbed the replisome (fig. 2.16(C)). 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 (fig. 2.16(D)). When proteins have been targeted for degradation, they are sent to another macromolecular assembly known as the proteasome (fig. 2.16(E)). The production of ATP in mitochondria is similarly mediated by a macromolecular complex known as ATP synthase (fig. 2.16(F)). 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

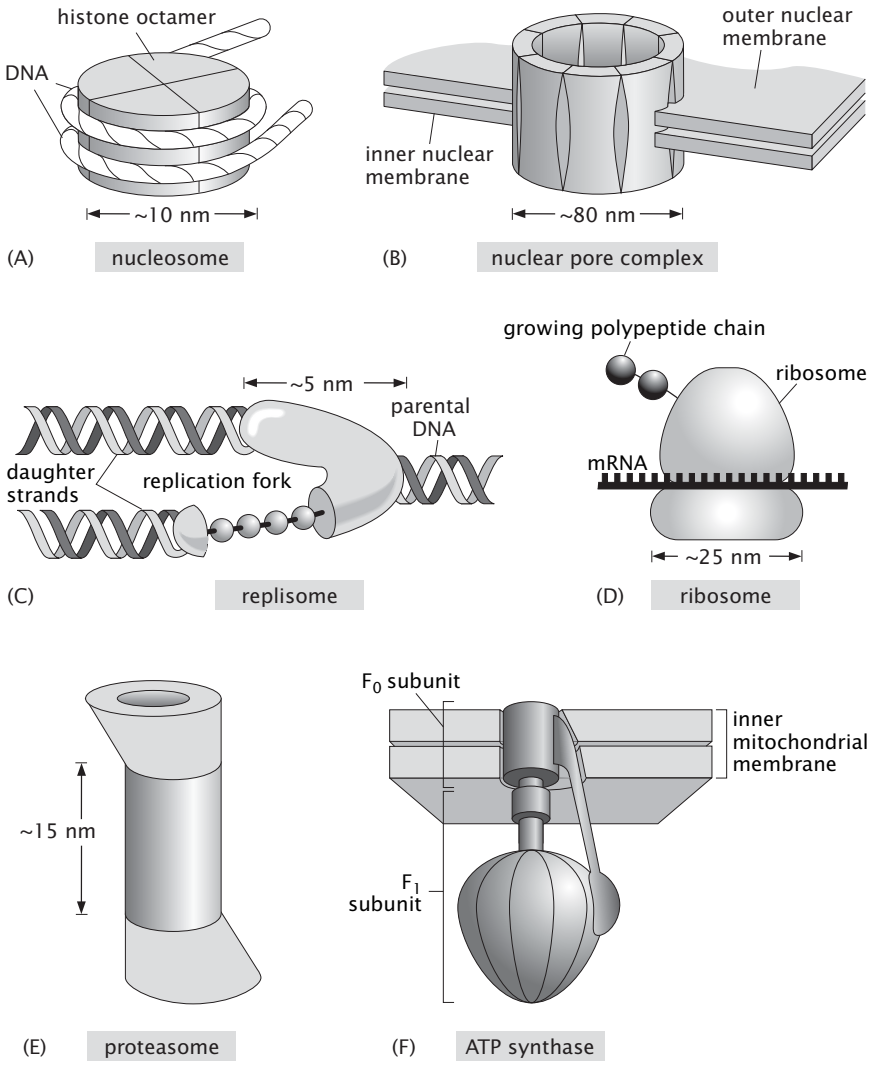

Figure 2.16: The macromolecular assemblies of the cell.

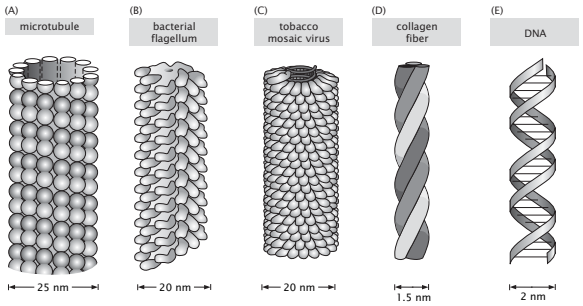

Figure 2.17: Helical motifs of molecular assemblies. Macromolecular assemblies have a variety of different helical structures, some formed from individual monomeric units (such as (A)-(C)) and others resulting from coils of proteins (D) and yet others made up of paired nucleotides (E).

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 shown in fig. 2.17. In fig. 2.17(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 fig. 2.17(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 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 of viruses.

The helical assemblies described above are characterized by individual protein units which come together to form helical filaments. An alternative and equally remarkable class of filaments are those in which alpha helices (chains of amino acids forming protein subunits with a precise, helical geometry) wind around each other to form superhelices. The particular case study which 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 fig. 2.8 on pg. 72). 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. Fig. 2.18 shows several examples. Fig. 2.18(A) shows the way in which ribosomes are organized on the endoplasmic reticulum with a characteristic spacing which is comparable to the size of the ribosomes ($\approx$ 20 nm). A second stunning example is the organization of myofibrils in muscles as shown in fig. 2.18(B). This figure shows the juxtaposition of the myofibrils and mitochondria. The

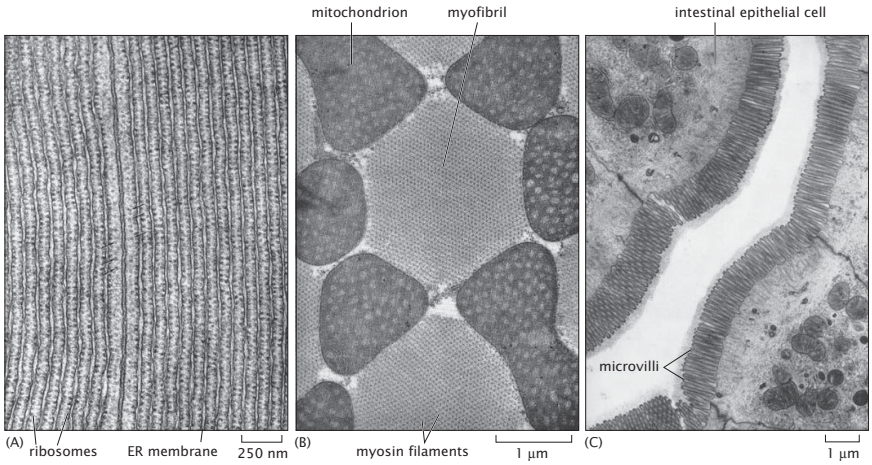

Figure 2.18: Ordered macromolecular assemblies. Collage of examples of macromolecules organized into superstructures. (A) ribosomes on the endoplasmic reticulum ("rough ER"), (B) myofibrils in the flight muscle, (C) microvilli at the epithelial surface. (A-C, adapted from D. W. Fawcett, The Cell: An Atlas of Fine Structure, Philadelphia: W. B. Saunders & Co., 1966.)

myofibrils themselves are an ordered arrangement of actin filaments and myosin motors as will be discussed in more detail in chap. 16. The last example shown in fig. 2.18(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 can 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 classes 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 10s to 100s of nanometers. The architecture of these 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. Fig. 2.19 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 these viruses is that by hijacking the host cell, the viral genome commands

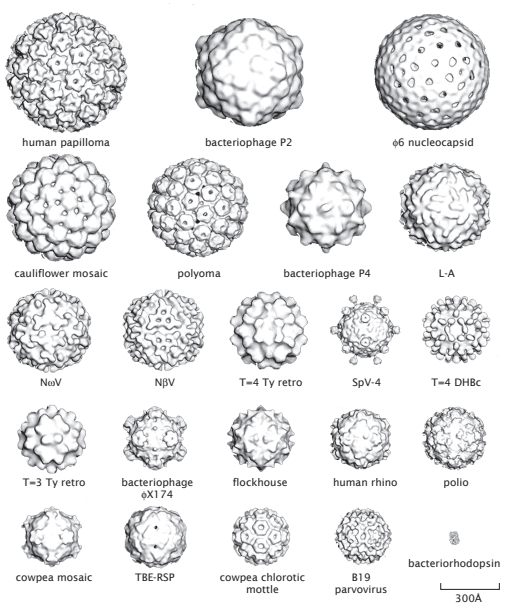

Figure 2.19: Structures of viral capsids. The regularity of the structure of viruses has enabled detailed, atomic-level analysis of their construction patterns. This gallery shows a variety of the different geometries explored by the class of nearly spherical viruses. For size comparison, a large protein bacteriorhodopsin is shown in the bottom right. (Adapted from T. S. Baker *et al.*, Microbiol. Mol. Biol. Rev., 63:862, 1999.)

the construction of its own inventory of parts within the host and then in the crowded environment of that host, assembles into infectious agents prepared to repeat the life cycle elsewhere.

HIV (human immunodeficiency virus) is one of the viruses that has garnered the most attention in recent years. Fig. 2.20 shows cryo-EM 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 kilo basepair (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

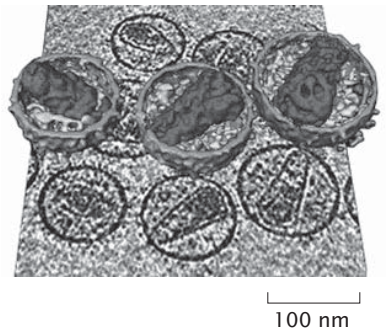

Figure 2.20: Structure of HIV viruses. The planar image shows a single frame from an electron microscopy tilt series. The three-dimensional images show reconstructions of the mature viruses featuring the ice-cream cone shaped capsid on the interior. (Adapted from J. A. G. Briggs *et al.*, Structure, 14:15, 2006.)

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 fig. 2.21. The formation of the *immature* virus occurs through the association of the N-terminal ends of these Gag proteins with the lipid bilayer of the host cell and 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 fig. 2.21. 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.

- **Estimate: Sizing Up HIV.** Unlike many of their more ordered viral counterparts, HIV virions have the intriguing feature that the structure from one to the next is not exactly the same. Indeed, they come in both different shapes and sizes. As a result, our attempt to "size up" HIV will be built around some representative numbers for these viruses, but the reader is cautioned to think of a statistical distribution of sizes and shapes. As shown in the cryo-EM picture of fig. 2.20, the size of the virion is between 120 nm and 150 nm and we take a "canonical" size of 130 nm.

  We begin with the immature virion. To find the number of Gag proteins within a given virion, we resort to simple geometrical reasoning. Since the

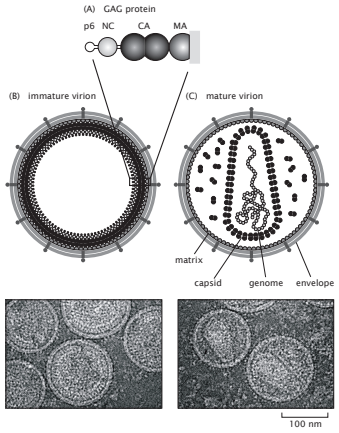

Figure 2.21: HIV architecture. (A) Schematic of the Gag polyprotein, a 41,000 Da architectural building block. (B) Immature virions showing the lipid bilayer coat and the uncut Gag shell on the interior, (C) mature virions in which the Gag protein has been cut by proteases and the separate components 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. Struc. Mol. Biol.,11:672, 2004.)

radius of the overall virion is roughly 65 nm, and the outer 5 nm of that radius is associated with the lipid bilayer, we imagine a sphere of radius 60 nm that is decorated on the inside with the inward facing spokes of the Gag proteins. If we think of each such Gag protein as a cylinder of radius 2 nm, this means they take up an area $A_{Gag} \approx 4\pi$ nm^2. Using this, we can find the number of such Gag proteins as

$$N_{Gag} = \frac{\text{surface area of virion}}{\text{area per Gag protein}} \approx \frac{4\pi(60 \text{ nm})^2}{4\pi \text{ nm}^2} \approx 3500. \tag{2.15}$$

The total mass of these Gag proteins is roughly

$$M_{Gag} \approx 3500 \times 40,000 \text{ Da} \approx 150 \text{ MDa}, \tag{2.16}$$

where we have used the fact that the mass of each Gag polyprotein is roughly 40 kDa. This estimate for the number of Gag proteins is of precisely the same magnitude as those that have emerged from recent cryo-electron microscopy observations (see Briggs, 2004).

The number of lipids associated with the HIV envelope can be estimated similarly as

$$N_{lipids} \approx \frac{2 \times 4\pi(65 \text{ nm})^2}{1/2 \text{ nm}^2} \approx 200,000 \text{ lipids}, \tag{2.17}$$

where the factor of 2 accounts for the fact that the lipids form a bilayer, and we have used a typical area per lipid of $1/2$ nm^2. The lipid census of HIV has been taken using mass spectrometry which permits the measurement of each of the different types of lipids forming the viral envelope (see Brügger, 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 classes 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 PDB Files to Ribbon Diagrams

If we continue with another factor of ten in our powers of ten 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 chap. 1 (pg. 25), these four classes of molecules make up the stuff of life and have central status in making up 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 one and ten nanometers. 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 $1/2$ 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 have 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 too. 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 chap. 8), at the level of structural biology, 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, fig. 1.1 (pg. 26) 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

$$\mathbf{r}_i = x_i\mathbf{i} + y_i\mathbf{j} + z_i\mathbf{k}, \tag{2.18}$$

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 eqn. 2.18 are really of the form

$$\mathbf{r}_i(t) = x_i(t)\mathbf{i} + y_i(t)\mathbf{j} + z_i(t)\mathbf{k}, \tag{2.19}$$

where the t reminds us that the coordinates depend upon time and what is measured in experiments might be best represented as $\langle \mathbf{r}_i(t) \rangle_{time}$, where the brackets $\langle \rangle_{time}$ 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 fig. 2.22. We choose this example because

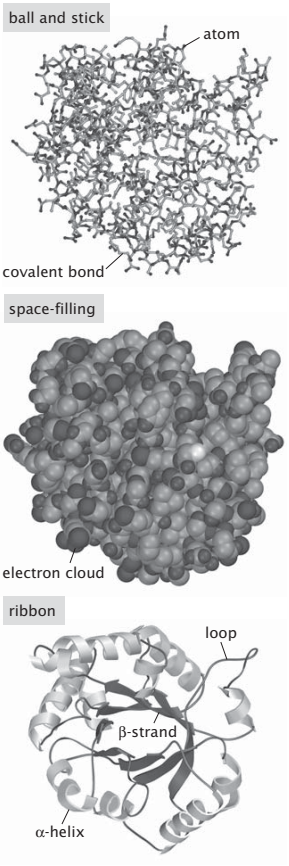

Figure 2.22: Three representations of triose phosphate isomerase. This enzyme is one of the enzymes in the glycolysis pathway. (Courtesy of David Goodsell.)

glycolysis will arise repeatedly throughout the book (see pg. 245 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 alpha helices and beta 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 figs. 1.1 and 2.22 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 which each have their own distinct identity as shown in fig. 2.23. The physical and chemical properties of these amino acids dictate both the folded shape of the protein as well as how it functions.

Amino acids are but one example of a broader class of 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 have their own unique chemical identity. Fig. 2.24 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 fig. 2.24 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 functional groups (serine, threonine, tyrosine and cysteine) are frequently important enzyme residues in catalytic reactions. The next row starts with a nitrogen containing amino group which is usually postively charged at neutral pH and a negatively charged carboxylic acid. 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 sidechains 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. On the other hand, these 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 fig. 1.3 (pg. 29), we showed the way in which individual groups can be seen as the building blocks of DNA structures such as that shown in fig. 1.1(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 which 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 ten descent which examine 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.

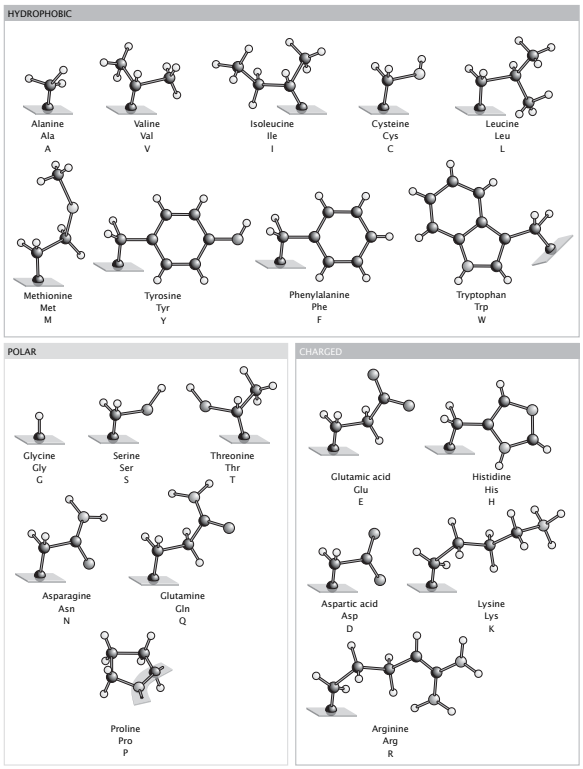

Figure 2.23: Amino acid side chains. The amino acids are represented here in ball and stick form, where a black ball indicates a carbon atom, a small white ball indicates a hydrogen atom and a gray ball, oxygen, nitrogen or sulfur. Only the side chains are shown. The peptide backbone of the protein to which these sidechains are attached is indicated by a flat gray 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.

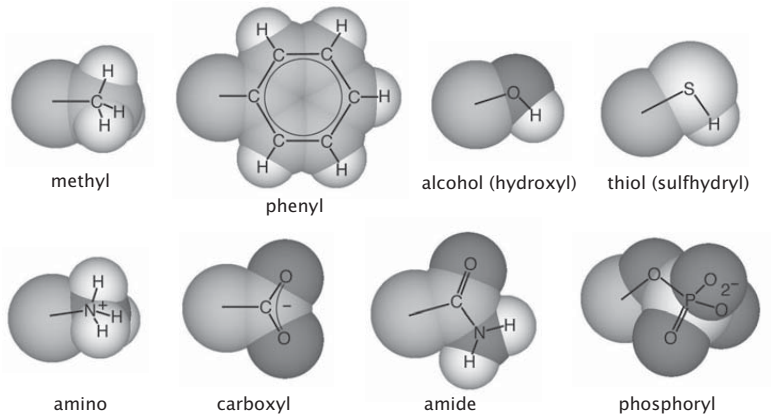

Figure 2.24: Chemical groups. These are some of the most common groups found in organic molecules such as proteins. (Courtesy of David Goodsell.)

2.3 Telescoping Up in Scale: Cells Don't Go It Alone

Our powers of ten journey has thus far shown us the way in which cells are built from structural units going down from organelles to macromolecular assemblies to individual macromolecules to chemical groups, atoms and ions. Equally interesting hierarchies of structures are revealed as we reduce the resolution of our imaginary camera and zoom out from the scale of individual cells. What we see once we begin to zoom out from the scale of single cells is the emergence of communities in which cells do not act independently.

2.3.1 Multicellularity As One of Evolution's Great Inventions

Life has been marked by several different evolutionary events which wrought a wholesale change in the way that cells operate. One important category of such events is the acquisition of the ability of cells to communicate and cooperate with one another to form multicellular communities with common goals. This has happened many times throughout all branches of life and has culminated in extremely large organisms such as redwood trees and giraffes among eukaryotes. In this section, we explore the ways in which cell-cell communication and cooperation have given rise to new classes of biological structures. Fig. 2.25 shows a variety of different examples of cellular communities, some of which form the substance of the remainder of the chapter.

Bacteria Interact to Form Colonies Such as Biofilms

The oldest known cellular communities recorded in the fossil record are es-

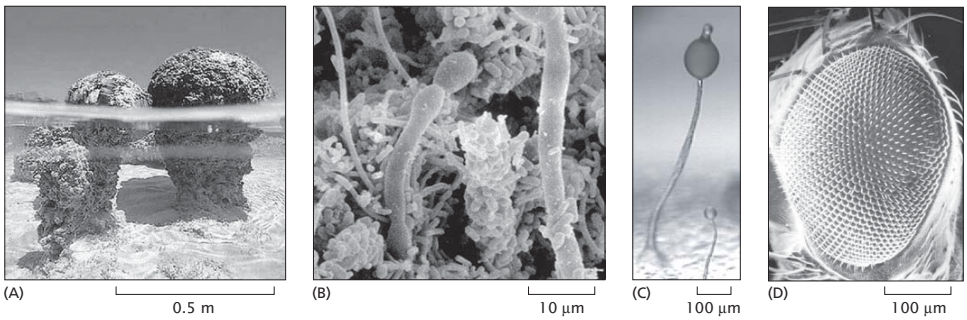

Figure 2.25: Representative examples of different communities of cells. (A) stromatolites from Australia. (B) Bacterial biofilm. (C) The social amoeba *Dictyostelium discoideum* forms fruiting bodies. The picture shows a fruiting body with spores - the tall stalk with a bulb at the top is a collection of amoebae. (D) The *Drosophila* eye. (A, adapted from *http://cas.bellarmine.edu/tietjen/Evolution/stromatolites2.htm*; B, adapted from *http://www.microbelibrary.org/ASMOnly/details.asp?id=541&Lang=English 061h.jpg*; C, adapted from *http://www.dundee.ac.uk/lifesciences/ps/organism.htm*; D, adapted from *http://8e.devbio.com/image.php?id=529*.)

sentially gigantic bacterial colonies called stromatolites. Although most stromatolites were outcompeted in their ecological niches by subsequent fancier forms of multicellular life, a few can still be found today taking essentially the same form as their two-billion year old fossils. Bacterial colonies of this type can still be found today in Australia such as those shown in fig. 2.25(A). These communities have a characteristic size of a meter and reflect collections of bacterial cells held together by an extracellular matrix secreted by these cells.

Many interesting kinds of bacterial communities consist of more than one species. Indeed, through a sophisticated system of signaling, detection and organization, bacteria form colonies of all kinds ranging from biofilms to the ecosystems within animal guts. Bacterial biofilms are familiar to us all as the basis of the dentist's warning to floss our teeth every night. These communities are functionally as well as structurally interdependent. Other biofilms are noted for their destructive force when they attach to the surfaces of materials. Fig. 2.25(B) shows a biofilm that grew on a silicon rubber voice prosthesis that had been implanted in a patient for about three months.

Structurally, a biofilm is formed as shown schematically in fig. 2.26. The key building blocks of such structures are a population of bacteria, a surface onto which these cells may adhere and an aqueous environment. The formation of a biofilm results in a population of bacterial cells that are attached to a surface and enclosed in a polymeric matrix built up of molecules produced by these very same bacteria. The early stages of biofilm formation involve the adhesion of the bacteria to a surface followed by changes in the characteristics of these bacteria

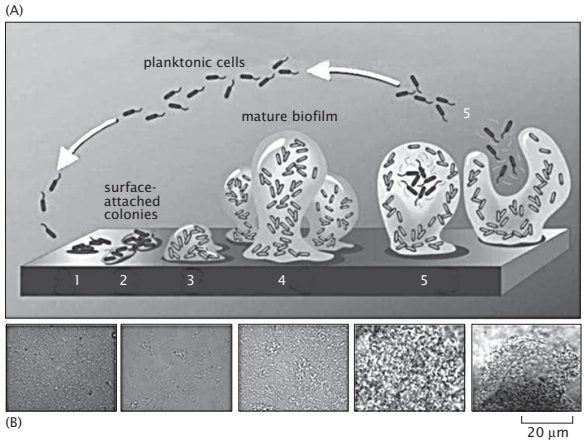

Figure 2.26: Schematic of the formation of a biofilm by bacteria. The various stages in the formation of the biofilm are: 1) attachment to surface, 2) secretion of extracellular polymeric substance (EPS), 3) early development, 4) maturation and 5) shedding of cells from the biofilm. The microscopy images below show biofilms in various stages of the film formation process. (Adapted from P. Stoodley *et al.*, Ann. Rev. Microbiol., 56:187, 2002.)

such as the loss of flagella and the development of pili. At a larger scale, these changes at the cellular level are attended by the formation of colonies of cells and differentiation of the colonies into structures which are embedded in extracellular polysaccharides. Though there are a variety of different morphologies that are adopted by such films, roughly speaking, these biofilms are relatively porous structures (presumably to provide a conduit for import and export of nutrients and waste, respectively) that typically take on mushroom-like structures such as indicated schematically in fig. 2.26. These films have a relative proportion of something like 85 percent of the mass taken up by extracellular matrix while the remaining 15 percent is taken up by cells themselves. A typical thickness for such films ranges from 10-50 μm.

Teaming Up in a Crisis: Lifestyle of *Dictyostelium discoideum*

Although bacteria can form communities, eukaryotes have clearly raised this to a high art. One particularly fascinating example that may give clues as to the origin of eukaryotic multicellularity is the cellular slime mold *Dictyostelium discoideum* as shown in fig. 2.25(C). This small, soil-dwelling amoeba pursues a solitary life when times are good but seeks the comfort of its fellows during times of starvation. *Dictyostelium* is usually content to wander around as an individual with a characteristic size between 5 and 10 μm. However, when deprived of its bacterial diet, these cells undertake a radical change in lifestyle which involves both interaction and differentiation through a series of fascinating

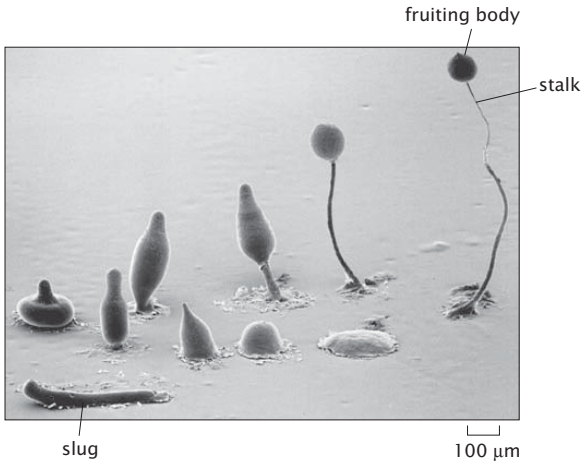

Figure 2.27: Formation of a multicellular structure during starvation. The social amoeba *Dictyostelium discoideum* responds to starvation by forming a structure made up of tens of thousands of cells in which individual cells suffer different fates. Cells near the top of the structure form spores which are resurrected once conditions are favorable. The figure shows the developmental stages that take place on the way to making fruiting bodies, starting in the bottom right and proceeding clockwise. (Copyright, M.J. Grimson and R.L. Blanton, Biological Sciences Electron Microscopy Laboratory, Texas Tech University)

intermediate steps as shown in fig. 2.27. A group of *Dictyostelium* amoebae in a soil sample that find themselves faced with starvation, send chemical signals to one another resulting in the coalescence of thousands of separate amoebae to form a slug that looks like a small nematode worm. These cells appear to be poised on the brink between unicellular and multicellular lifestyles and can readily convert between them. Ultimately, as shown in the figure, the slug stops moving and begins to form a stalk. At the tip of the stalk is a nearly spherical bulb that contains many thousands of spores, essentially cells in a state of suspended animation. When environmental conditions are appropriate for individual amoeba to thrive, the spores undergo the process of sporulation, with each spore becoming a functional, individual amoeba.

- **Estimate: Sizing Up the Slug and the Fruiting Body.** The relation between the number of cells in a slug and its size is shown in fig. 2.28 where we see that these slugs can range in size between several hundred and several thousand microns with the number of cells making up the slug between tens of thousands and several million. The next visible stage in the development is the sprouting of a stalk with a bulb at the top known as a fruiting body. The stalk is on the order of a millimeter in length while the size of the fruiting body itself is several hundred microns

across. This fruiting body is composed of thousands of spores, amoeba that are effectively in a state of suspended animation. An example of such a fruiting body that has been squished on a microscope slide is shown in fig. 2.29. This structure has functional consequences. In particular, those cells that are part of the spore remain poised to respond to a better day, while the cells that formed the stalk have effectively ended their lives for the good of those that survive.

An immediate question of interest concerning the multicellular fruiting bodies shown in fig. 2.27 is how many cells conspire to make up such structures. Fig. 2.29 provides the answer, but it is also of interest to try to reason it out. An estimate of the number of cells in a fruiting body can be constructed by examining the nearly hemispherical collection of cells shown in fig. 2.29. Note that we treat the fruiting body as a hemisphere since once it is squashed on the microscope slide, its shape is flattened and resembles a hemisphere more closely than a sphere. The diameter of this hemisphere is roughly = 200 μm. Our rough estimate for the number of spores in the fruiting body is obtained by evaluating the ratio

$$\text{number of cells} = \frac{V_{body}}{V_{cell}}, \tag{2.20}$$

where we assume that the entirety of the fruiting body volume is made up of cells. If we assume that the cell size is 4 μm in diameter, this yields

$$\text{number of cells} = \frac{\frac{2}{3}\pi(100\ \mu m)^3}{\frac{4}{3}\pi(2\ \mu m)^3} \approx 6 \times 10^4 \text{cells}. \tag{2.21}$$

Note that the size of the ball of cells in a fruiting body can vary dramatically from the 200 μm scale shown here to several times larger, and as a result, the estimate for the number of cells in a fruiting body can vary. Also, note that a factor of two error in our estimate of the size of an individual cell will translate into a factor of eight error in our count of the number of cells in the slug or fully formed fruiting body.

Multicellular Organisms Have Many Distinct Communities of Cells

The three branches of life that have most notably exploited the potential of the multicellular lifestyle are animals, plants and fungi. While bacterial and protozoan colonial organisms rarely form communities with characteristic dimensions of more than a few millimeters (with the exception of stromatolites), individuals in these three groups routinely grow to more than a meter in size. Their enormous size and corresponding complexity can be attributed to at least three factors: i) production of extracellular matrix material that can provide structural support for large communities of cells, ii) a predilection towards cellular specialization such that many copies of cells with the same genomic content can develop to perform distinct functions and iii) highly sophisticated mechanisms for the cells to communicate with one another within the organism. We

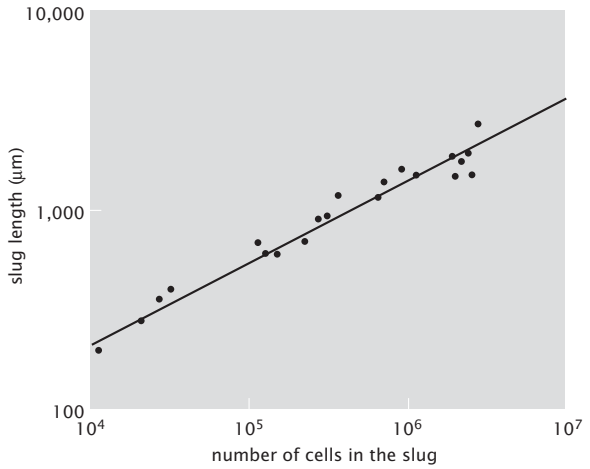

Figure 2.28: Slug size in *Dictyostelium discoideum*. The plot shows a relation between the size of the Dicty slug and the corresponding number of cells. (adapted from Bonner, www.dictybase.org)

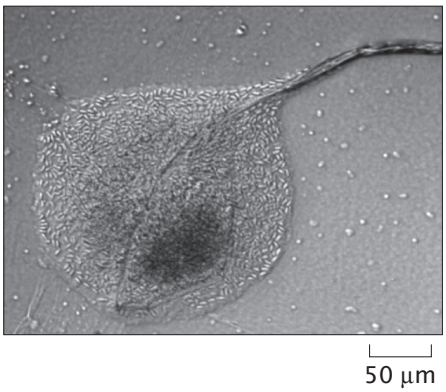

Figure 2.29: Microscopy image of a fruiting body. The fruiting body has been squished on a microscope slide revealing both the size of the spores and their numbers.

emphasize that these traits are not unique to animals plants and fungi, but they are used more extensively there than elsewhere. A beautiful illustration of these principles is seen in the eye of the fruit fly *Drosophila* as shown in fig. 2.25(D). The eye is made up of hundreds of small units called ommatidia, each of which contains a group of eight photoreceptor cells, support cells and a cornea. During development of the eye, these cells signal to one another to establish their identities and relative positions to create a stereotyped structure that is repeated many times. The overarching theme of the remainder of this chapter is the exploration of how cells come together to form higher order structures and how these structures fit into the overall hierarchy of structures formed by living organisms.

2.3.2 Cellular Structures From Tissues to Nerve Networks

Multicellular structures are as diverse as cells themselves. Often, the nature of these structures are a reflection of their underlying function. For example, the role of epithelia as barriers dictates their tightly packed, planar geometries. By way of contrast, the informational role of the network of neurons dictates an entirely different type of multicellular structure.

One Class of Multicellular Structures Is the Epithelial Sheets

Epithelial sheets form part of the structural backdrop in organs ranging from the skin to the bladder. Functionally, the cells in these structures have roles such as serving as a barrier to transport of molecules, providing an interface at which molecules can be absorbed into cells and as the seat of certain molecular secretions. Several different views of these structures are shown in fig. 2.30.

The morphology of epithelial sheets are diverse in several ways. First, the morphology of the individual cells can be different (isotropic vs anisotropic). In addition, the assemblies of cells themselves have different shapes. For example, the different structures can be broadly classified into those which are a monolayer sheet (simple epithelium) and those which are a multilayer (stratified epithelium). Within these two broad classes of structures, the cells themselves have different morphologies. The cells making up a given epithelial sheet can be flat, pancake-like cells, denoted as *squamous* epithelium. If the cells making up the epithelial sheet have no preferred orientation, they are referred to as *cuboidal*, while those which are elongated perpendicular to the extracellular support matrix are known as *columnar* epithelia. Epithelial sheets have as one of their functions (as do lipid bilayers) the segregation of different media which can have highly different ionic concentrations, pH, macromolecular concentrations and so on.

Tissues Are Collections of Cells and Extracellular Matrix

We have seen that cells can interact to form complexes. An even more intriguing example of a multicellular structure is provided by tissues such as that shown in fig. 2.31. These connective tissues are built up from a diverse array of cells and materials they secrete. Beneath the epithelial surface, fibroblasts

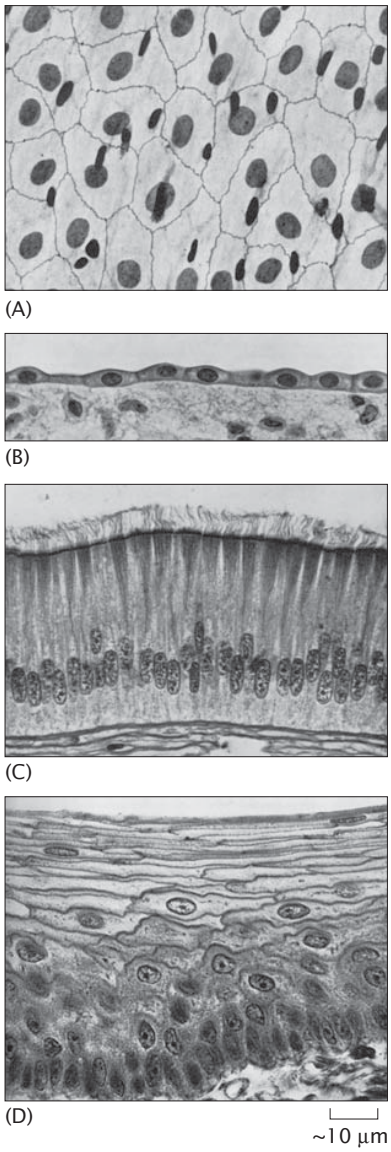

Figure 2.30: Shapes and architecture of epithelial cells. Epithelia are tissues formed by continuous sheets of cells that form tight contacts with one another. (A) Viewed from above, a simple epithelial sheet resembles a tiled mosaic. The dark ovals are the cell nuclei stained with silver. (B) Viewed from the side, simple epithelia such as this from the dog kidney, may form a single layer of flattened cells above a loose, fibrous connective tissue. (C) In some specialized epithelia, the cells may extend upwards forming elongated columns and develop functional specializations at the top surface such as the beating cilia shown here. This particular ciliated columnar epithelium is from the alimentary tract of a freshwater mussel. (D) In other epithelia such as skin or the kitten's gum shown here, epithelial cells may form multiple layers. (A-D, adapted from D. W. Fawcett, The Cell: An Atlas of Fine Structure, Philadelphia: W. B. Saunders & Co., 1966)

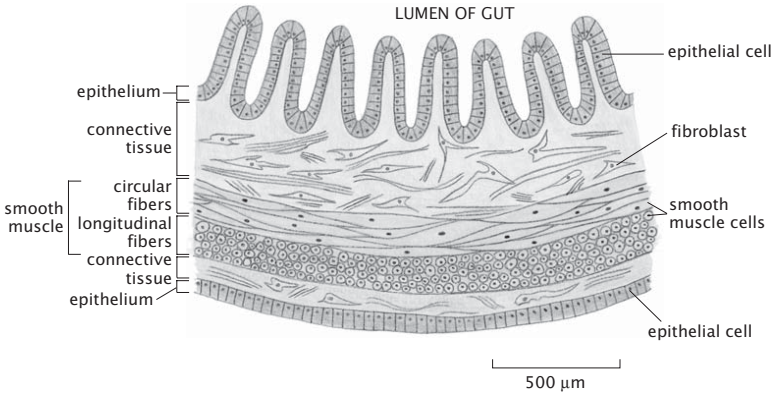

Figure 2.31: Connective tissue. The schematic shows the organization of cells and extracellular matrix that make up connective tissue. The top layer is a planar array of epithelial cells. The volume beneath these cells is made up of fibroblasts and a secretion of extracellular matrix. (Adapted from B. Alberts *et al.*, Molecular Biology of the Cell, 4th ed. New York: Garland Science, 2002.)

construct extracellular matrix. This connective tissue is built up of three main components: cells such as fibroblasts and macrophages, connective fibers and a structureless supporting substance made up of glycosaminoglycans (GAGs). What is especially appealing and intriguing about these tissues is the orchestration, both in space and time, leading to the positioning of cells and fibers. The fibroblasts, which were already featured earlier in this chapter, serve as factories for the proteins that make up the extracellular matrix. In particular, they synthesize proteins such as collagen and elastin which, when secreted, assemble to form fibrous structures which can support mechanical loads. The medium within which these fibers (and the cells) are embedded is made up of the third key component of the extracellular matrix, namely, the hydrated gel of glycosaminoglycans.

Nerve Cells Form Complex, Multicellular Complexes

A totally different example of structural organization involving multiple cells is illustrated by collections of neurons. Neurons are the specialized cells in animals that we associate with thinking and feeling. These cells allow for the transmission of information over long distances in the form of electrical signals. Neurons are constructed with many input terminals known as dendrites and a single output path known as the axon. The fascinating structural feature of these cells is that they assemble into complex networks that are densely connected in patterns where dendrites from a given cell reach out to many others, from which they take various inputs. An example of a collection of fluorescently labeled neurons is shown in fig. 2.32. Note that the branches (dendrites and axons) that reach out from the various cells have lengths far in excess of the 10

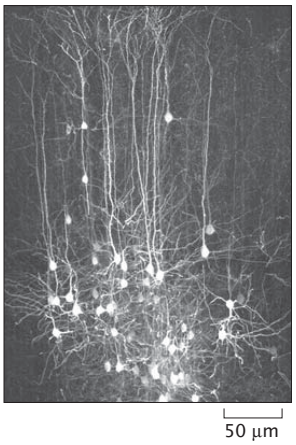

Figure 2.32: Illustration of a complex network of cells formed by neurons. GFP fluorescence is used to label a collection of neurons from the brain of a rat. Particular neurons were targeted using lentiviruses. (Adapted from M. Brecht *et al.*, J. Neurosci., 24:9233, 2004.)

micron scale characteristic of typical eukaryotes. Indeed, axons of some neurons can have lengths of centimeters and more.

One fascinating example of neuronal contact is offered by the so-called neuromuscular junction as shown in fig. 2.33. These junctions are the point of contact between motor neurons, which convey the marching orders for a given muscle, and the muscle fiber itself. As is seen in the figure, the axon from a given motor neuron makes contacts with various muscle fibers. As will be described in more detail in chap. 17, when an electrical signal (action potential) arrives at the contact point known as a synapse, chemical neurotransmitters are released into the space between the nerve and muscle. These neurotransmitters result in the opening of ligand-gated ion channels in the muscle which result in a change in the electrical state of the muscle and lead to motion of the muscle. Contrasting the contacts between epithelial sheets and neurons (or neurons and muscles) reveals the diversity of cell-cell contacts.

2.3.3 Multicellular Organisms

The highest level in the structural hierarchy to be entertained here is individual organisms. The diversity of multicellular organisms is legendary ranging from roses to hummingbirds, Venus flytraps to the giant squid. What is especially remarkable about this diversity is contained in the simple statement of the cell theory: each and every one of these organisms is a collection of cells and their products. And, each and every one of these organisms is the result of a long history of evolution resulting in specialization and diversification of the various

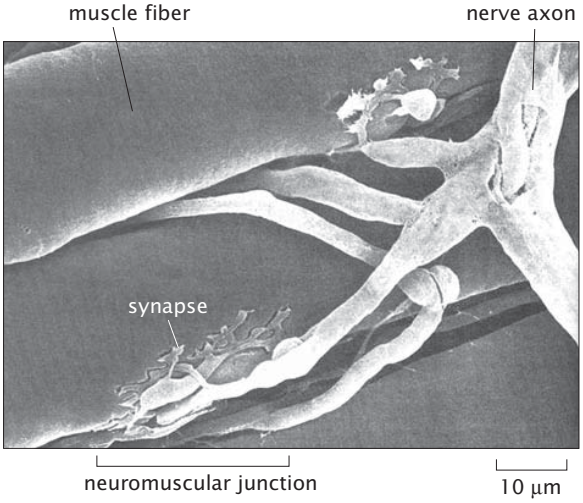

Figure 2.33: Neuromuscular junction. The axon from a nerve cell makes contact with various skeletal muscle fibers. Each muscle fiber is a giant multinucleated cell formed by the fusion of hundreds or thousands of precursor cells. Neurotransmitters secreted by the nerve cell at the synapse initiate contraction of the muscle fibers. (Adapted from D. W. Fawcett and R. P. Jensh, Bloom and Fawcett's Concise Histology, London, Arnold Publishers, 1997.)

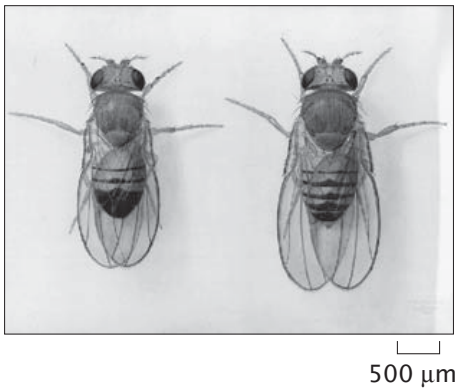

Figure 2.34: The male and female adult wild-type *Drosophila melanogaster*. After 1934 drawing by E. M. Wallace. (Courtesy of the Archives, California Institute of Technology)

cells that make up that organism.

Cells Differentiate During Development Leading to Entire Organisms

The fruit fly *Drosophila melanogaster* has had a long and rich history as one of the key "model" organisms of biology and is a useful starting point for thinking about the size of organisms. As shown in fig. 2.34, the mature fly has a size of roughly 3 mm that can be thought of morphologically as being built up of 14 segments: 3 segments making up the head, 3 segments making up the thorax and 8 segments making up the abdomen. We will further explore both the life cycle and the biological significance of *Drosophila* in chaps. 3 (pg. 156) and 4 (pg. 218), respectively, and for now content ourselves with examining the spatial scales associated with both the mature fly (fig. 2.34) and the embryos from which they are derived.

Drosophila has attained its legendary status in part because of the way it has revealed so many different ideas about embryonic development. One of the most well-studied features of the development of *Drosophila* is the way in which it lays down its anterior-posterior architecture during early development. This body plan is dictated by different cells adopting different patterns of gene expression depending upon where they are within the embryo. The pattern of expression of the so-called even-skipped gene is shown in fig. 2.35. The gene *even-skipped* (*eve*) is expressed in seven stripes corresponding to seven of the fourteen *Drosophila* segments. One of the most remarkable features of the pattern of gene expression illustrated in the figure is its spatial sharpness with the discrimination between the relevant gene being "on" or "off" taking place on a length scale comparable to the cell size itself. It is intriguing that the gene expression profile at the level of individual cells in the embryo correspond to macroscopic territories in the mature animal.

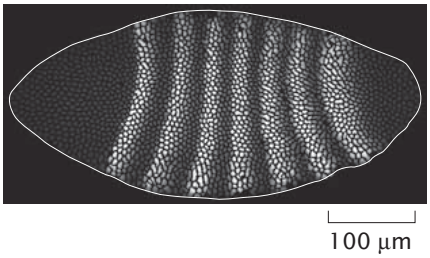

Figure 2.35: Pattern of gene expression in the *Drosophila* embryo. Image of *Drosophila* embryo with immunofluorescence staining using anti-Eve antibody. Each spot corresponds to the nucleus of a different cell making up the football shaped embryo. This embryo has undergone 13 cycles of cell division. (Adapted from E. Poustelnikova *et al.*, Bioinformatics, 20:2212, 2004.)

- **Estimate: Sizing Up Stripes in *Drosophila* Embryos.** The sharp patterns in gene expression exhibited by the pair-rule genes such as the *even-skipped* gene shown in the figure result from spatial gradients of DNA binding proteins known as transcription factors that either increase (activation) or decrease (repression) the level of gene expression. To get a feeling for the scales associated with the gradients in transcription factors that dictate developmental decisions and the features they engender, we idealize the *Drosophila* embryo as a spherocylinder. We characterize the geometry by two parameters, the length of the cylindrical region, L, and its radius R, where we use approximate values of 300 μm for the length of the cylindrical region and 100 μm for the radius. In the embryo, many of the nuclei are on the surface and as a result, our estimates will depend upon having a rough estimate of their areal density. The area of the embryo in the simple geometrical model used here is given by

$$A = 4\pi R^2 + 2\pi RL. \tag{2.22}$$

  If we consider the embryo after 13 divisions (during so-called stage 14) and before the gastrulation which folds the developing embryo in, there are roughly 6000 nuclei on the surface of the embryo. Note that there are not 2^{13} nuclei since during the preceding divisions, not all of the nuclei went to the surface of the embryo, and those that are left behind in the interior of the embryo do not keep pace with those on the surface during subsequent divisions. See Zalokar and Erk (1976) for details of the nuclear census during embryonic development. The areal density (i.e. number per unit area) is given by

$$\sigma = \frac{N}{4\pi R^2 + 2\pi RL}. \tag{2.23}$$

Using the numbers described above leads to an areal density of 0.02 cells $/\mu\text{m}^2$. As seen in fig. 2.35, the stripe patterns associated with the *Drosophila* embryo are very sharp and reflect cells making decisions at a very localized level.

Inspection of fig. 2.35 reveals that the stripes are roughly 20 μm wide. As a result, we can estimate the total number of cells participating in each stripe as

$$n = \sigma 2\pi R l_{stripe} \approx 0.02 \text{ cells}/\mu\text{m}^2 \times 2 \times 3.14 \times 100 \ \mu\text{m} \times 20 \ \mu m \approx 250. \tag{2.24}$$

Note that the average area per cell is given by $1/\sigma \approx 50 \ \mu\text{m}^2$, suggesting that the radii of these cells is roughly 4 μm. Our main purpose in carrying out this exercise is to demonstrate the length scale of the structures that are put down during embryonic development. In this case, what we have seen is that out of the roughly 6000 cells that characterize the *Drosophila* embryo at the time of gastrulation, groups of roughly 250 cells have begun to follow distinct pathways as a result of differential patterns of gene expression which foreshadow the heterogeneous anatomical features of the mature fly.

The Cells of the Nematode Worm *C. elegans* Have Been Charted Yielding a Cell By Cell Picture of the Organism

A more recently popularized model organism for studying the genetics and development of multicellular animals is the nematode worm *Caenorhabditis elegans* shown in fig. 2.36. Several factors that make this worm particularly attractive in its capacity as a model multicellular eukaryote are that a) its complete genome has been determined, b) the identity of each and every of its 959 cells has been determined (see fig. 2.37) and c) the worms are transparent permitting a variety of structural investigations. Amazingly, *all* of the cells of this organism have had their lineages traced from the single ancestral cell which is present at the moment of fertilization. What this means precisely is that all of the roughly 1000 cells making up this organism can be assigned a lineage of the kind cell A begat cell B which begat cell C and so on.

As shown in fig. 2.36, these worms are roughly 1 mm in length and 0.05 mm across. Like *Drosophila*, they too have been subjected to a vast array of different analyses including, for example, how their behavior is driven by the sensation of touch. One of the most remarkable outcomes of the series of experiments leading to the lineage tree for *C. elegans* was the determination of the connectivity of the 302 neurons present in this nematode. By using serial thin sections from electron microscopy, it was possible to map out the roughly 7000 neuronal connections in the nervous system of this tiny organism. The various nerve cells are typically less than 5 microns across.

- **Estimate: Sizing Up *C. Elegans*.** As a simple estimate of the cellular content of a "simple" organism, *Caenorhabditis elegans* provides a way

Figure 2.36: *C. elegans*. This DIC image of a single adult worm was assembled from a series of high-resolution micrographs. The worm's head is at the top left corner and its tail is at the right. Its gut is visible as a long tube going down the animal's body axis. Its egg cells are also visible as giant ovals towards the bottom of the body. (Courtesy of Ian D. Chin-Sang)

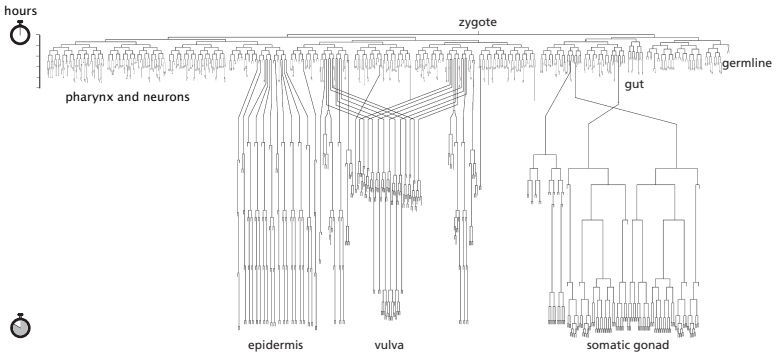

Figure 2.37: Cell lineages in *C. elegans*. The developmental pattern of every cell in the worm is identical from one animal to the other. Therefore, it has been possible for developmental biologists to determine the family tree of every cell present in the entire animal by patient direct observation. In this schematic representation, the cell divisions that occur during embryogenesis are shown in the band across the top. The later cell divisions of the epidermis, vulva and somatic gonad all take place after the animal has hatched. The Xs represent cells that die; death is a normal developmental fate.

of estimating the characteristic size of eukaryotic cells. As suggested by fig. 2.36, these small worms can be thought of as cylinders of length 1 mm and with a width of 0 .05 mm. The total volume of such a worm is computed simply as roughly 5×10^6 μm^3. If we use the fact that there are roughly 1000 cells in the organism, this tells us that the characteristic volume of the cells is 5000 μm^3, implying a typical radius for these cells of roughly 10 μm. A reasonable estimate for the volume of eukaryotic cells is somewhere between 2000 - 10000 μm^3.

Many different species of nematodes share a common body plan involving a small number of well defined cells, however they may vary greatly in size. One of the largest nematodes *Ascaris*, is a common parasite of pigs and humans. It has been estimated that up to one billion people on the planet carry *Ascaris* in their intestine. These worms closely resemble *C. elegans* except that they may be up to 15 cm in length. This kind of observation is not unusual throughout the animal kingdom. Many species have close relatives that differ enormously in size. The reasons and mechanisms that determine the overall size of multicellular organisms remain poorly understood.

Higher Level Structures Exist as Colonies of Organisms

Organisms do not exist in isolation. Every organism on the planet is part of a larger ecological web that features both cooperation and competition with other individuals and species. Coral reefs represent a vivid example of the interdependence of huge varieties of species living together in close quarters. An equally vivid and diverse community, though one frequently less appreciated is found closer to home within our own intestines, which are inhabited by a teeming variety of bacteria. It has been estimated that the human body actually harbors nearly ten times more microbial cells than it does human cells, and at least one hundred times more bacterial genes than human genes. Thus we should think of ourselves not really as individuals but rather as complex ecosystems of which the human cells form only a small part. Thus, our story of the hierarchy of structures that make up the living world began with the bacterium *Escherichia coli* and will end there now. It is a great irony that we as humans, it might be argued, have been colonized by our standard ruler *E. coli*.

Though we have examined biological structures over a wide range of spatial scales, our powers of ten journey still falls far short of being comprehensive. Ultimately, what we learn from this is that the handiwork of evolution has resulted in biological structures ranging from the nanometer scale of molecular machines all the way to the scale of the planet itself. Trying to understand what physical and biological factors drive the formation of these structures will animate much of the remainder of the book.

2.4 Summary and Conclusions

Biological structures range from the scale of nanometers (individual proteins) to hundreds of meters (redwood trees) and beyond (ecological communities).

In this chapter, we have explored the sizes and numbers of biological entities starting with the unit of life, the individual cell and working our way down and up. We have found that sometimes biological objects show up in very large numbers of identical copies such that averages, for example, concentration of ribosomes in the *E. coli* cytoplasm, are reasonable approximations. On the other hand, sometimes biological things show up in very few copies such that the exact number can make a big difference in the behavior of the system. We have found that cells are crowded, that there really is a world in a grain of sand (or in a biological cell). We have also explored some of the ways that biological units such as proteins or cells may interact with one another to form larger and more complex entities. Having developed an intuitive feeling for size and scale will better enable us to realistically envision the biological processes and problems described throughout the remainder of the book.

2.5 Problems

1. A feeling for the numbers: microbes as the unseen majority.

(a) Use fig. 2.1 (pg. 58) to justify the assumption that a typical bacterial cell (i.e. *E. coli*) has a surface area of 6 μm^2 and a volume of 1 μm^3. Also, express this volume in femtoliters. Make a corresponding estimate of the mass of such a bacterium.

(b) Roughly 2-3 kg of bacteria are harbored in your large intestine. Make an estimate of the total number of bacteria inhabiting your intestine. Estimate the total number of human cells in your body and compare the two figures.

(c) The claim is made (see Whitman *et al.* (1998)) that in the top 200 m of the world's oceans, there are roughly 10^{28} prokaryotes. Work out the total volume taken up by these cells in m^3 and km^3. Compute their mean spacing. How many such cells are there per ml of ocean water?

2. A feeling for the numbers: molecular volumes and masses.

(a) Estimate the volumes of the side chains of the various amino acids in nm^3 units.
(b) Estimate the mass of a "typical" amino acid in Daltons. Justify your estimate by explaining how many of each type of atom you chose. Compare your estimate to several key amino acids such as glycine, proline, arginine and tryptophan.
(c) On the basis of your result for part (b), deduce a rule of thumb for converting mass of a protein (reported in kDa) into a corresponding number of residues. Apply this rule of thumb to myosin, G-actin, hemoglobin and hexokinase and compare your result to the actual number of residues in each of these proteins. Relevant data for this problem is provided on the book website.

3. Minimal media and *E. coli*.

Minimal growth medium for bacteria such as *E. coli* includes various salts with characteristic concentrations of mM and a carbon source. This carbon source is typically glucose and it is used at 0.2 % (a concentration of 0.2 g/100 ml).
(a) Make an estimate of the number of carbon atoms it takes to make up the macromolecular contents of a bacterium such as *E. coli*.
(b) How many cells can be grown in a 5 mL culture using minimal media before the medium exhausts the carbon? Note that this estimate will be flawed because it neglects the *energy* cost of synthesizing the macromolecules of the cell. These shortcomings will be addressed in chap. 5.

4. Atomic-Level representations of biological molecules.

(a) Obtain coordinates for several of the following molecules: ATP, phosphatidylcholine, B-DNA, G-actin, the lambda repressor/DNA complex or Lac repressor/DNA complex, hemoglobin, myoglobin, HIV gp120, green fluorescent protein (GFP) and RNA polymerase. You can find the coordinates on the book website or by searching in the Protein Data Bank and various other Internet resources.

(b) Download a structural viewing code such as VMD (University of Illinois), Rasmol (University of Massachusetts) or DeepView (Swiss Institute of Bioinformatics) and create a plot of each of the molecules you downloaded above. Experiment with the orientation of the molecule.

(c) By looking at phosphatidylcholine justify (or improve upon) the value of the area per lipid (0.5 nm^2) used in the chapter.

(d) Phosphoglycerate kinase is a key enzyme in the glycolysis pathway. One intriguing feature of such enzymes is their enormity in comparison with the sizes of the molecules upon which they act (their "substrate"). This statement is made clear in fig. 5.5 (pg. 250). Obtain the coordinates for both phosphoglycerate kinase and glucose and examine the relative size of these molecules. The coordinates are provided on the book website.

5. HIV estimates.

(a) Estimate the total mass of an HIV virion by comparing its volume to that of an *E. coli* cell and assuming they have the same density.

(b) The HIV maturation process involves proteolytic clipping of the Gag polyprotein so that the capsid protein CA can form the shell surrounding the RNA genome and nucleocapsid NC can complex with the RNA itself. Using figs. 2.20 (pg. 90) and 2.21 (pg. 91) to obtain the capsid dimensions, estimate the number of CA proteins that are used to make the capsid and compare your result to the total number of Gag proteins.

6. Areas and volumes of organelles.

(a) Calculate the average volume and surface area of mitochondria in yeast based on the confocal microscopy image of fig. 2.9(D) (pg. 75).
(b) Estimate the area of the ER when it is in reticular form using a model for its structure of interpenetrating cylinders of diameter $d \approx 10$ nm separated by a distance $a \approx 60$ nm, as shown in fig. 2.15 (pg. 83).

7. An open ended "feeling for the numbers": the cell.

Use the book by Fawcett (1966) to examine various pictures of cells and their organelles and make up your own series of estimates. Some possible examples include: the spacing between lamellae in mitochondria, the volume of sperm heads, the dimensions of the parts in muscle cells, etc. (pg. 83).

2.6 Further Reading

D. Goodsell, **The Machinery of Life**, Springer-Verlag, New York: New York, 1998. This book illustrates the crowded interior of the cell and the various macromolecules of life.

D. W. Fawcett, **The Cell, Its Organelles and Inclusions**, W. B. Saunders and Company, Philadelphia, Pennsylvania, 1966. We imagine our reader comfortably seated with a copy of Fawcett right at his or her side. Fawcett's electron microscopy images are stunning.

D. W. Fawcett and R. P. Jensh, **Bloom and Fawcett's Concise Histology**, Arnold Publishers, London, England, 1997. This book shows some of the beautiful diversity of both cells and their organelles.

S. F. Gilbert, **Developmental Biology**, Sinauer Associates, Sunderland: Massachusetts, 2003. Gilbert's book is a source for learning about the architecture of a host of different organisms during early development. Chap. 9 is especially relevant for the discussion of this chapter.

L. Wolpert L., R. Beddington, T. Jessell, P. Lawrence, E. Meyerowitz, and J. Smith, **Principles of Development**, Oxford University Press, New York: New York, 2002. Wolpert's book is full of useful cartoons and schematics that illustrate many of the key ideas of developmental biology.

2.7 References

T. S. Baker, N. H. Olson and S. D. Fuller, "Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from

cryo-electron micrographs", Microbiol Mol Biol Rev. **63**, 862, (1999).

M. Brecht, M. S. Fee, O. Garaschuk, F. Helmchen, T. W. Margrie, K. Svoboda and P. Osten, "Novel Approaches to Monitor and Manipulate Single Neurons *in vivo*", J. Neurosci., **24**(2), 9233 (2004).

J. A. G. Briggs, M. N. Simon, I. Gross, H.-G. Kraüsslich S. D. Fuller, V. M. Vogt, M. C. Johnson, "The stoichiometry of Gag protein in HIV-1", Nat. Struc. Mol. Biol., **11**, 672 (2004).

J. A. G. Briggs, K. Grünewald, B. Glass, F. Förster, H.-G. Kraüsslich and S. D. Fuller, "The Mechanism of HIV-1 Core Assembly: Insights from Three-Dimensional Reconstruction of Authentic Virions", Structure, **14**, 15 (2006).

B. Brügger, B. Glass, P. Haberkant, I. Leibrecht, F. T. Wieland and H.-G. Kraüsslich, "The HIV lipidome: a raft with an unusual composition", Proc. Nat. Acad. Sci. **103**, 2641 (2006).

D. J. Chin, K. L. Luskey, R. G. W. Anderson, J. R. Faust, J. L. Goldstein and M. S. Brown, "Appearance of crystalloid endoplasmic reticulum in compactin-resistant Chinese hamster cells with a 500-fold increase in 3-hydroxy-3-methylglutaryl-coenzyme A reductase", Proc. Nat. Acad. Sci., **79**, 1185 (1982).

P. C. Cross and K. L. Mercer, **Cell and Tissue Ultrastructure**, W. H. Freeman and Company, New York: New York, 1993. A great pictorial source for anyone interested in the beauty of cells and their organelles.

D. W. Fawcett, **The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure**, W. B. Saunders and Company, Philadelphia: Pennsylvania, 1966.

D. W. Fawcett, **Bloom and Fawcett - A Textbook of Histology**, W. B. Saunders Company, Philadelphia: Pennsylvania, 1986.

S. Ghaemmaghami, W.-K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea and J. S. .Weissman, "Global analysis of protein expression in yeast", Nature **425**, 737 (2003).

F. C. Neidhardt, J. L. Ingraham and M. Schaechter, **Physiology of the Bacterial Cell**, Sinauer Associates, Inc., Sunderland: Massachusetts, 1990. This outstanding book is full of interesting facts and insights concerning bacteria.

S. Pedersen, P. L. Bloch, S. Reeh and F. C. Neidhardt, "Patterns of Protein Synthesis in E. coli: a Catalog of the Amount of 140 Individual Proteins at Different Growth Rates", Cell **14**, 179 (1978).

E. Poustelnikova, A. Pisarev, M. Blagov, M. Samsonova and J. Reinitz, "A database for management of gene expression data in situ", Bioinformatics, **20**, 2212-2221 (2004).

M. Schaechter, J. L. Ingraham and F. C. Neidhardt, **Microbe**, ASM Press, Washington DC, 2006.

P. Stoodley, K. Sauer, D. G. Davies and J. W. Costerton, "Biofilms as Complex Differentiated Communities", Ann. Rev. Microbiol. **56**, 187 (2002).

M. Terasaki, L. B. Chen and K. Fujiwara, "Microtubules and the Endoplasmic Reticulum Are Highly Interdependent Structures", J. Cell. Biol., **103**, 1557 (1986).

W. Visser, E. A. Van Sponsen, N. Nanninga, J. T. Pronk, J. G. Kuenen and J. P. van Dijken, "Effects of growth conditions on mitochondrial morphology in *Saccharomyces cerevisiae*", Antonie van Leeuwenhoek **67**, 243 (1995).

G. M. Walker, **Yeast, Physiology and Biotechnology**, John Wiley and Sons, Chichester: England, 1998.

W. B. Whitman, D. C. Coleman and W. J. Wiebe, "Prokaryotes: the unseen majority", Proc. Natl. Acad. Sci. **95**, 12, 6578 (1998).

J.-Q. Wu and T. D. Pollard, "Counting Cytokinesis Proteins Globally and Locally in Fission Yeast", Science **310**, 310 (2005). This paper illustrates the strategy of using fluorescence microscopy for taking the census of a cell.

W. Yu, J. M. Solowska, L. Qiang, A. Karabay, D. Baird and P. W. Baas, "Regulation of microtubule severing by katanin subunits during neuronal development", J Neurosci. **25** 5573, (2005).

M. Zalokar and I. Erk, "Division and Migration of Nuclei during Early Embryogenesis of *Drosophila melanogaster*", J. Microscopie Biol. Cell. **25**, 97 (1976). This paper makes a census of the number of cells in the *Drosophila* embryo at various stages in development.

S. B. Zimmerman and S. O. Trach, "Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of *Escherichia Coli*", J. Mol. Biol., **222**, 599 (1991). Discussion of crowding in the bacterial cell.