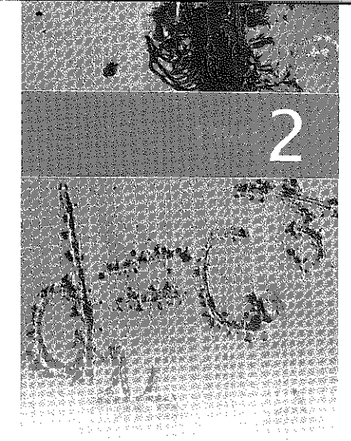


2 What and Where: Construction Plans for Cells and Organisms

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* (commonly abbreviated as *E. coli*) 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. 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-10 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 the biological structures 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.

> "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**

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, in contrast to single-celled organisms.

Cells are able to consume energy from their environments and use that energy to create ordered structures. They can also harness energy and materials 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 (generating 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 cell 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 (or, more accurately, a community of common ancestors) that would be recognized as cellular by any modern biologist.

In terms of understanding the basic rules governing metabolism, replication, and life more generally, one cell type should be as good as any other as the basis of experimental investigations of these mechanisms. 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 have 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. Second, 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 to make DNA and RNA 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 of the consequence of mutations 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 among many cells or even among different species. 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.

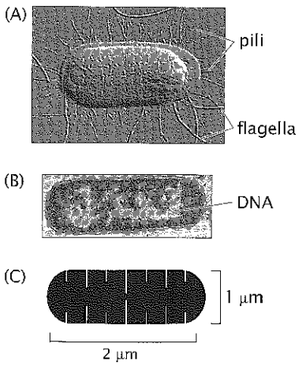


Figure 2.1 *E. coli* as a standard ruler for characterizing spatial scales. (A) Atomic force microscopy image of an *E. coli* cell, (B) electron micrograph of an *E. coli* bacterium, and (C) the *E. coli* ruler. (A, courtesy of C. T. Lim.)

Figure 2.1 shows several experimental pictures of an *E. coli* cell and its schematization into our standard ruler. The electron micrograph in Figure 2.1(B) shows that these bacteria have a rod-like morphology with a typical length between 1 and 2 μm and a diameter between 1/2 and 1 μm. To put the standard ruler in perspective, we note that with its characteristic length scale of 1 μm, it would take roughly 50 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 500 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 with which they are provided, with those growing in richer media having a larger size. Our reference growth condition throughout the book will be a chemically defined solution referred to by microbiologists as "minimal medium" that is a mixture of salts along with glucose as the sole carbon source. 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. For consistency, we will therefore refer primarily to experimental results for bacterial growth in minimal media.

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 experiments performed *in vitro* (that is, in test tubes) and *in vivo* (that is, in living cells). 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 surface 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 $V_{E.\ coli} \approx 1\,\mu m^3 = 1 fL$ to *E. coli* and an area of roughly

$A_{E.\ coli} \approx 6\,\mu m^2$ (see the problems for examples of how to 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 to 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 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 Chapter 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 Chapter 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 first to 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 intermediate 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, 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 with 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% 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 the order of 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 toward revealing the extent of crowding within a bacterium, we can estimate the number of proteins by assuming an average 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 $5 \times 10^{-20}\,g$. The number of proteins per *E. coli* cell is estimated as

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

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 the order of 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% 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 that: (a) the mass of an individual ribosome is roughly 2.5 MDa and (b) 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

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

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. Our estimate for that

part of the ribosomal mass that is protein is 830,000 Da. 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 $10^8\,\text{nm}^3$. This is 10% 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 $A_{E.\,coli} \approx 6\,\mu\text{m}^2$. 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 $0.5\,\text{nm}^2$. The measured number of lipids is of the order of 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 obtain 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

$$\begin{aligned} N_{H_2O} &\approx \frac{0.7 \times 10^{-12}\,\text{g}}{18\,\text{g/mol}} \times 6 \times 10^{23}\ \text{molecules/mol} \\ &\approx 2 \times 10^{10}\ \text{water molecules.} \end{aligned} \tag{2.4}$$

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

$$\begin{aligned} N_{ions} &\approx \frac{(100 \times 10^{-3}\ \text{mol}) \times (6 \times 10^{23}\ \text{molecules/mol})}{10^{15}\,\mu\text{m}^3} \times 1\,\mu\text{m}^3 \\ &= 6 \times 10^7. \end{aligned} \tag{2.5}$$

Here we use the fact that $1\,\text{L} = 10^{15}\,\mu\text{m}^3$. 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 Figure 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 with our estimates) in Table 2.1.

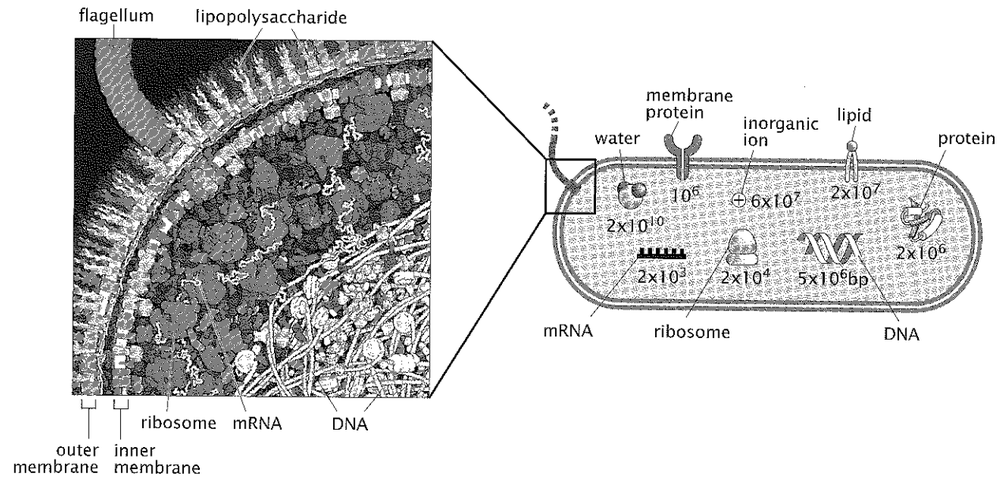


Figure 2.2 Molecular contents of the bacterium *E. coli*. The illustration on the left shows the crowded cytoplasm of the bacterial cell. The cartoon on the right shows an order-of-magnitude molecular census of the *E. coli* bacterium with the approximate number of different molecules in *E. coli*. (Illustration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

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 Figure 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 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

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, 2006.)

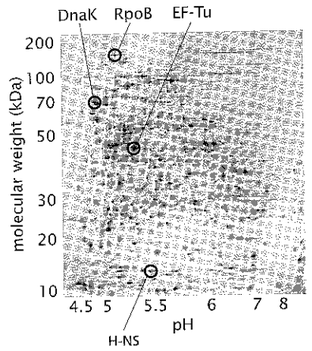


Figure 2.3 Protein census of the cell. Measurement of protein census of *E. coli* using two-dimensional polyacrylamide gel electrophoresis. Each spot represents an individual protein with a unique size and charge distribution. The spots arising from several well-known bacterial proteins are labeled. (Copyright Swiss Institute of Bioinformatics, Geneva, Switzerland "Copyright Swiss Institute of Bioinformatics, Geneva, Switzerland.")

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 left-hand corner and small, positively charged proteins at the lower right-hand corner for the gel shown in Figure 2.3. The proteins can then be stained with a nonspecific dye so that their 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 be determined by physically cutting each spot out of the gel, eluting the protein, and determining its size and amino acid content using mass spectrometry. 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.

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

One of the most intriguing implications of our census of the molecular parts 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 are shown in Figure 2.2. This figure gives a number of different views of the crowding associated with any *in vivo* process. In Chapter 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 Figure 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 a concentration of 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 Figure 2.4. This figure shows the number of copies of the molecule of interest in such a cell as a function of the concentration.

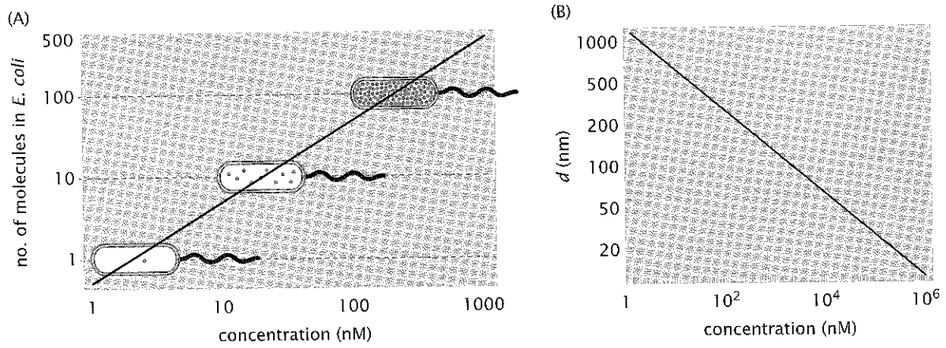


Figure 2.4 Physical interpretation of concentration. (A) Concentration in *E. coli* units: number of copies of a given molecule in a volume the size of an *E. coli* cell as a function of the concentration. (B) Concentration expressed in units of typical distance (d) between neighboring molecules measured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

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 Figure 2.4(B) which shows the relation between the mean spacing measured in nanometers and the concentration in nanomolar units.

2.1.3 Looking Inside 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 biological structures are explored and measured.

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 originated. 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.

Figure 2.5 shows three distinct experimental strategies which feed into our estimates, all of which reveal different facets of biological structure. One of the mainstays of structural analysis is light microscopy. Figure 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 fluorophores 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, it is only these structures that are observed when the cell is examined in the microscope. 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).

EXPERIMENTS

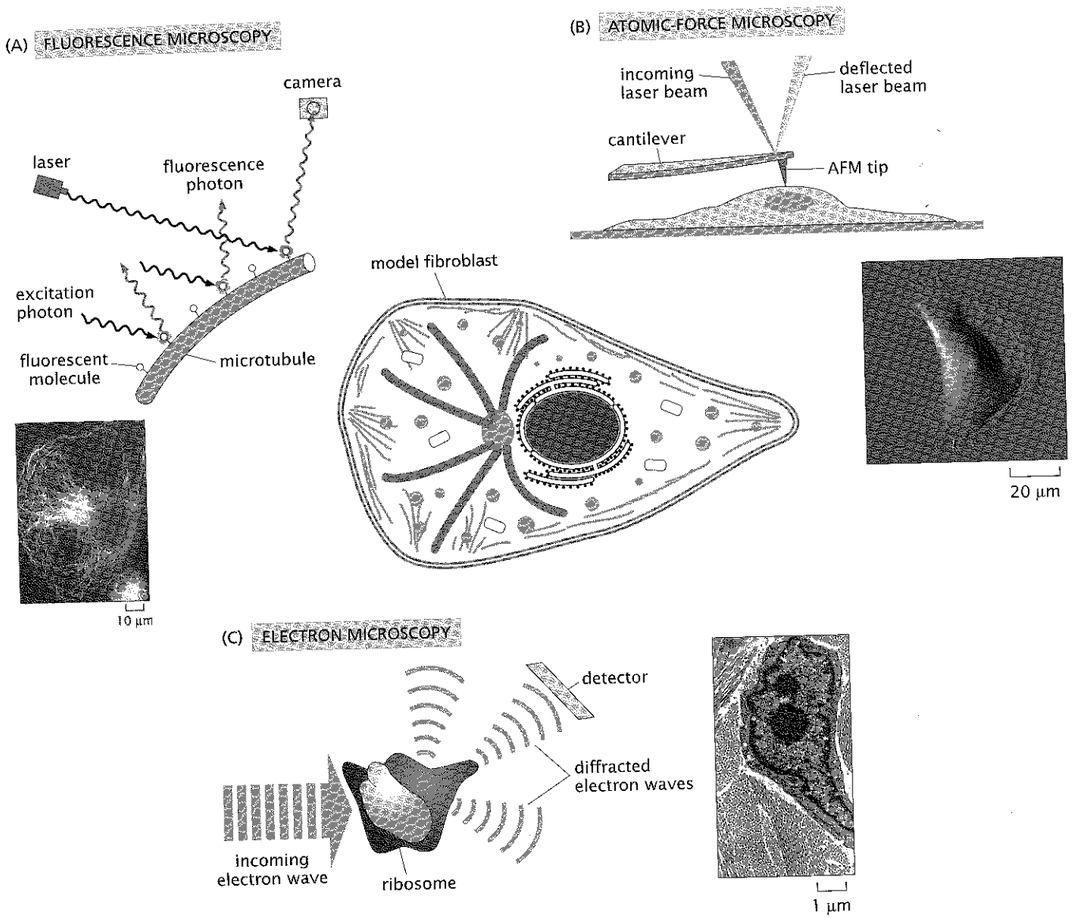


Figure 2.5 Experimental techniques which have revealed the structure of both cells and their organelles. (A) Fluorescence microscopy and the associated image of a fibroblast with labeled microtubules. (B) Atomic-force microscopy schematic and the associated image of the surface topography of a fibroblast. (C) Electron microscopy schematic and image of cross-section through a fibroblast in an animal tissue. (A, adapted from E. Schulze and M. Kirschner, *J. Cell Biol. 104*, 277, 1987; 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.)

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 Chapter 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. Figure 2.5(B) shows a schematic of an AFM scanning a typical fibroblast cell as well as a corresponding image of the cell.

Figure 2.5(C) gives a schematic of the way in which X-rays or electrons are scattered off 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

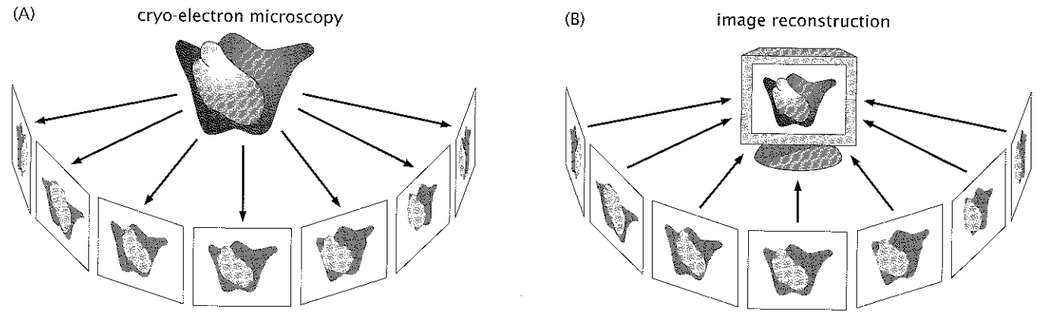


Figure 2.6 Schematic of tomographic reconstruction. (A) The sample is rotated so that radiation is scattered from a series of different orientations. (B) Three-dimensional reconstruction of the structure giving rise to the pattern of scattering.

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.

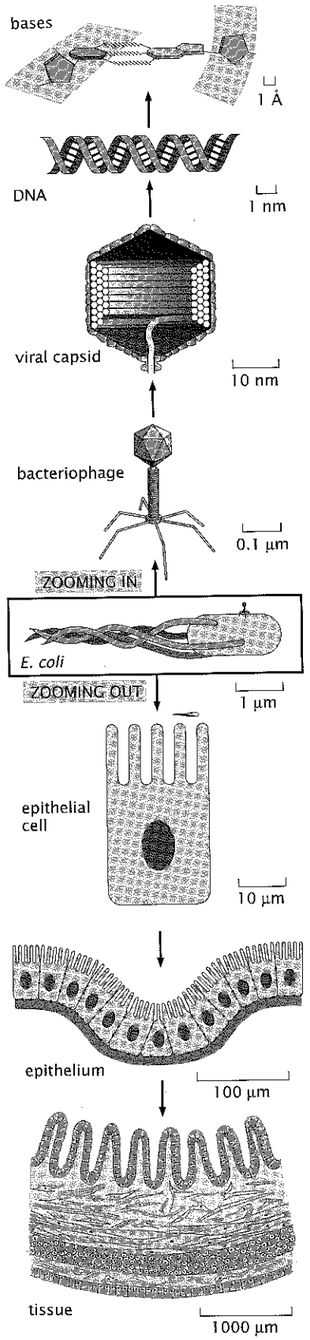
An important variation on the theme of measuring the scattered intensity from irradiated samples is cryo-electron tomography. This technique is one of the centerpieces of structural biology and is built around uniting electron microscopy with sample preparation techniques which rapidly freeze the sample. The use of tomographic methods has made it possible to go beyond the planar sections seen in conventional electron microscopy images. The basis of the technique is indicated schematically in Figure 2.6, and relies on rotating the sample over a wide range of orientations and then assembling a corresponding three-dimensional reconstruction on the basis of the entirety of these images. This technique has already revolutionized our understanding of particular organelles and is 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? Figure 2.7 shows the different structures that can be seen as we scale in and out from an *E. coli* cell. A roughly 10-fold increase in magnification relative to an individual bacterium reveals the viruses that attack bacteria. These viruses, known as bacteriophage, have a characteristic scale of 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 10 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 Figure 2.7, a 10-fold 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 central players as part of our intestinal ecosystem. Another 10-fold 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 and organs. Scaling out to larger scales would bring us to multicellular organisms and the structures they build.

The remainder of the chapter takes stock of the structures at each of these scales and provides 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 of which they are made. 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.

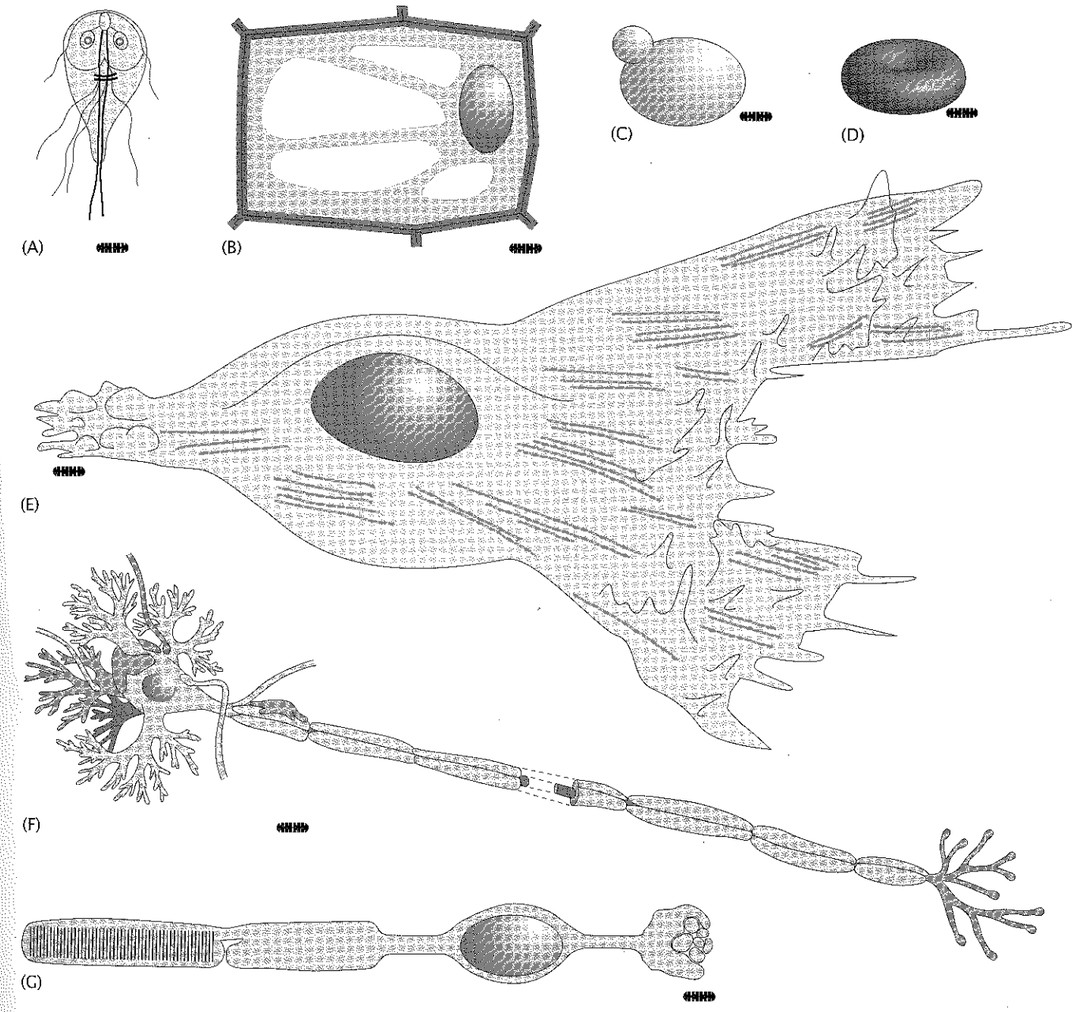
2.2 CELLS AND STRUCTURES WITHIN THEM

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 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 Eukarya.

The organisms that we most often 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.

Figure 2.7 Powers of 10 representation of biological length scales. The hierarchy of scales is built around the *E. coli* standard ruler. Starting with *E. coli*, Section 2.2 considers a succession of 10-fold increases in resolution as are shown in the figure. Section 2.3 zooms out from the scale of an *E. coli* cell.



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 Figure 2.8 we show a rogue's gallery illustrating a small segment of 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

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

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 plankton 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*. Figure 2.8(A) shows the intriguing protist *Giardia lamblia*, a parasite known to hikers as a source of water contamination that causes diarrhea. A broader set of examples of protists is shown in Figure 20.3 (p. 784).

Although protists constitute the vast majority of eukaryotic cells on the planet, biologists are often inclined to study cells more closely related to or directly useful to us. This includes the plant kingdom which is obviously important as a source of food and flowers. Plant cells like that shown in Figure 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 micrometers. One of the distinctive features of plant cells 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 an animal cell of similar overall size, the plant cell will have typically 10-fold 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 the eukaryotes, the group most closely related to the animals (as proved by ribosomal RNA similarity and many other lines of evidence) is, surprisingly, the fungi. The representative fungus shown in Figure 2.8 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* often 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. cerevisiae*. Therefore, many laboratory 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). Figure 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 μm. Nonetheless, it possesses all the important structural hallmarks of the eukaryotes including, in particular, a membrane-bound nucleus, separating the DNA genome from the cytoplasmic machinery that performs most metabolic functions.

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.

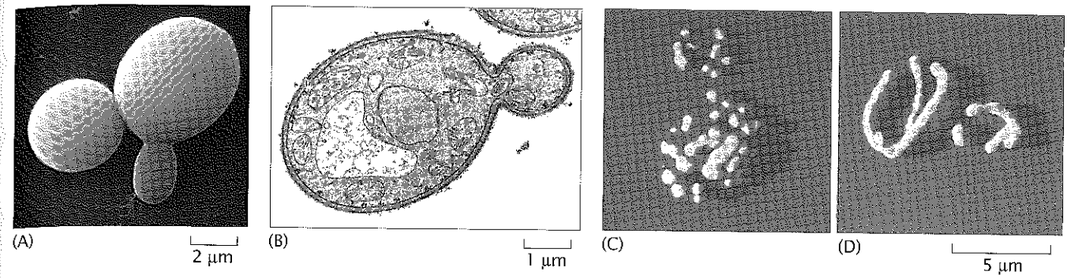


Figure 2.9 Microscopy images of yeast and their organelles. (A) Scanning electron micrograph of the yeast *Candida albicans* revealing the overall size scale of budding yeast. (B) Electron microscopy image of a cross-section through a budding *Candida albicans* yeast cell. (C) and (D) Confocal microscopy images of the mitochondria of *S. cerevisiae*. (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.)

ESTIMATE

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 $V_{E.\,coli} \approx 1.0\,\mu m^3$ 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 $4\,\mu m^3$. Within this nucleus is housed the 1.2×10^7 bp of the yeast genome which is divided among 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

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

Experimentally, the measured number appears to be closer to 80,000, with a mean spacing between nucleosomes of the order of 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 $14 \times 10^6\,nm^3$ taken up by the histones. The volume taken up by the genome itself is comparable at $1.2 \times 10^7\,nm^3$, 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 nucleus) associated with the yeast genomic DNA can be estimated by evaluating the ratio

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

Note that we have used the fact that the yeast genome is 1.2×10^7 bp in length and is packed in the nucleus which has a volume of $\approx 4\,\mu m^3$.

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 in 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

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

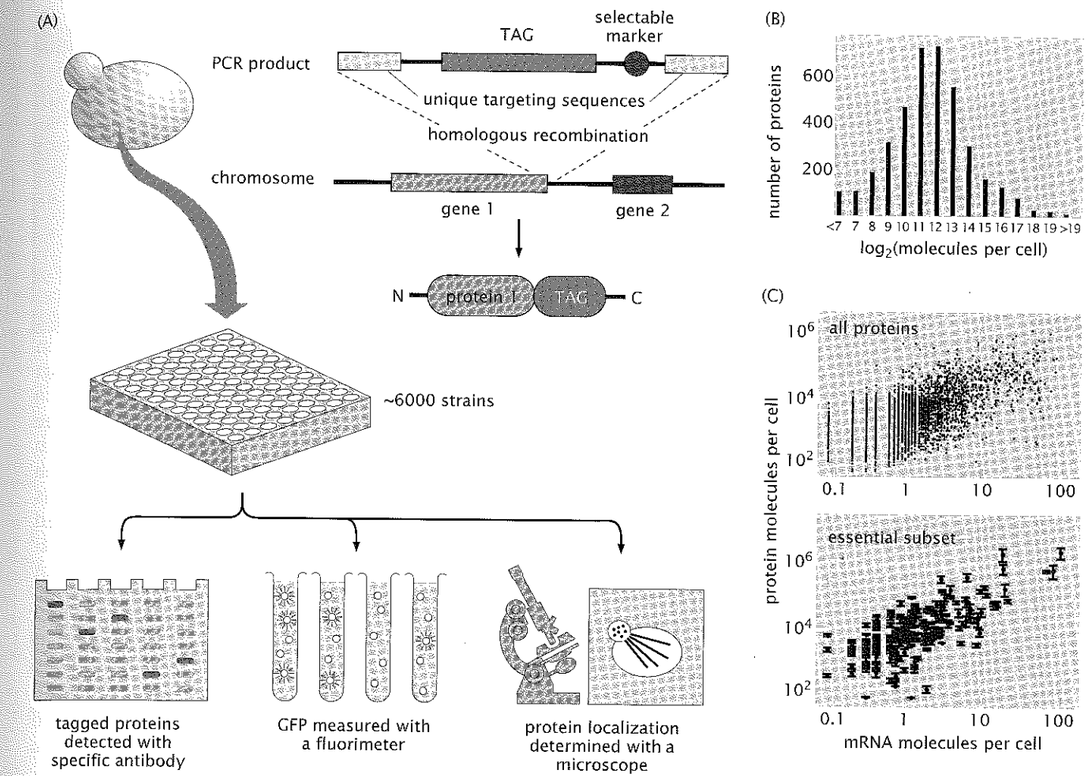
where the factor of 0.5 is based on the idea that roughly half of the surface area is covered by membrane proteins rather than lipids themselves and the factor of 2 accounts for the fact that the membrane is a bilayer. This should be contrasted with the situation in *E. coli* which has a double membrane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

Another interesting estimate suggested by Figure 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 in several ways – one of which is to attempt to count them directly, another 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 Figure 2.9(C). Further, we estimate that the typical mitochondrial size is roughly 3/4 μm, resulting in a total mitochondrial volume of

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

which given the total volume of the cell of $60\,\mu\mathrm{m}^3$ translates into a volume fraction of roughly 15%. The total area of the outer membranes of these mitochondria is roughly $70\,\mu\mathrm{m}^2$, comparable to the entire area of the plasma membrane itself. The analysis of the image in Figure 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. For a series of recent studies, 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 attached to each protein in the cell 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. Figure 2.10(B) shows a histogram of the number of proteins that occur with a given protein copy number in yeast. 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 50 copies and others with more than a million copies. Further, similarly quantifying mRNA as shown in Figure 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 Figure 2.8.

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. The "tag" attached to each gene may encode a fluorescent protein or a site for antibody recognition, depending on the experiment. (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.)

Cells from Humans Have a Huge Diversity of Structure and Function

The remainder of the cells in Figure 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