

2.1.1 The Bacterial Standard Ruler

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

Throughout this 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 shows several experimental pictures of an *E. coli* cell and its schematization into our standard ruler. The AFM image and the electron micrograph in Figures 2.1(A) and (B) show that these bacteria have a rod-like morphology with a typical length between 1 and 2 μm and a diameter between 0.5 and 1 μm . The reader is invited to use light microscopy images of these bacteria and to determine the size of an *E. coli* cell for him or herself in the “Computational Exploration” at the end of this section.

The length unit of 1 μm , or micrometer, is so useful for the discussion of cell biology that it has a nickname, the “micron.” To put this 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. Using these insights, the question of how many bacteria can dance on the head of a pin is answered unequivocally in Figure 2.1(D).

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 mass. An extremely elegant experiment that explores the connection between growth rate and mass is shown in Figure 2.2. 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

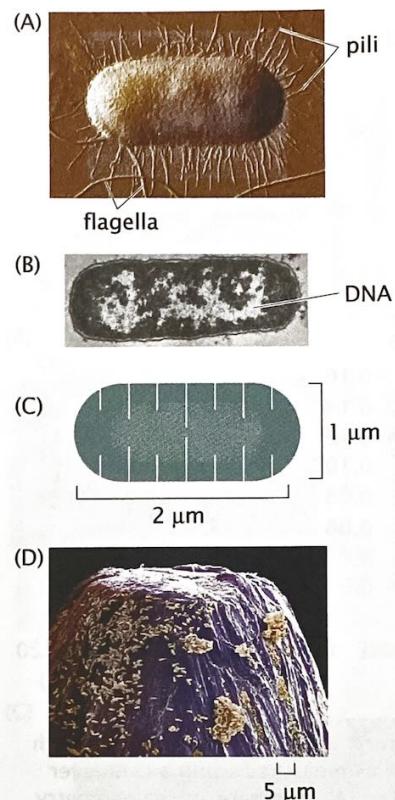


Figure 2.1: *E. coli* as a standard ruler for characterizing spatial scales. (A) Atomic-force microscopy (AFM) image of an *E. coli* cell, (B) electron micrograph of a sectioned *E. coli* bacterium, and (C) the *E. coli* ruler. (D) Bacteria on the head of a pin giving an impression of how our standard ruler compares with the dimensions of everyday objects. (A, courtesy of Ang Li; D, courtesy of Tony Brain.)

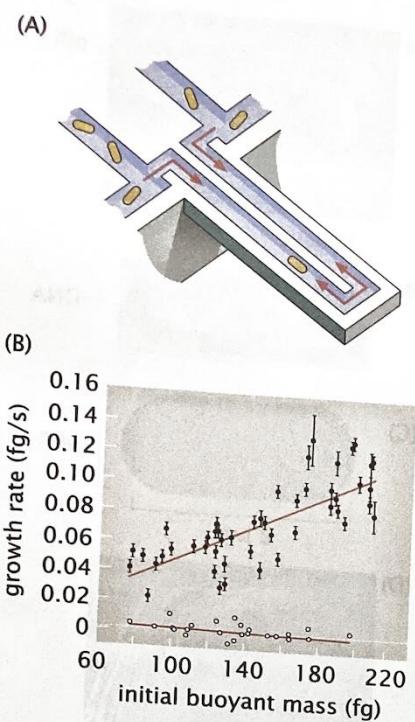


Figure 2.2: *E. coli* mass and growth rate as measured using a cantilever assay. (A) Schematic of the geometry of the device used to characterize the buoyant mass of bacterial cells by dynamically trapping them within the hollow cantilever. The hollowed out cantilever oscillates with a slightly different frequency when there is a cell present in its interior. (B) Relation between the buoyant mass at the start of the experiment and the growth rate. *E. coli* K12 cells grown at 37°C. Filled circles correspond to cells that are growing normally and open circles correspond to fixed cells. (Figures adapted from M. Godin et al., *Nat. Meth.* 7: 387, 2010.)

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 inside the cell (transcription factors are 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 $V_{E. coli} \approx 1 \mu\text{m}^3 = 1 \text{ fL}$ to *E. coli* and an area of roughly $A_{E. coli} \approx 6 \mu\text{m}^2$ (see the problems for examples of how to work out these numbers from known cellular dimensions).

Computational Exploration: Sizing Up *E. coli* As already revealed in Figure 2.1, the simplest way for us to figure out how big bacteria are is to stick them under a microscope and to look at them. Of course, to do this in reality requires some way to relate the cells we see magnified in our microscope and their actual size. In modern terms, what this really means is figuring out the relationship between the size of the pixels produced by our camera and real world length units such as microns.

In this Computational Exploration, the reader is invited to take the images shown in Figures 2.3(A) and (C) and to use them to determine how many nanometers are in a pixel. To that end, the idea is to read the image file from Figure 2.3(A) into Matlab and to produce a graph of the intensity for one row of the image as shown in Figure 2.3(B). Each one of the lines in the graticule shown in Figure 2.3(A) is separated by 10 μm. With this information in hand, the reader can work out the size of our pixels and hence install the missing scale bar on Figure 2.3(C), which is the ultimate objective of this Computational Exploration.

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. As we will see below this distance is comparable to the size of the proteins themselves!

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, though clearly Figure 2.2 shows that this assumption is not true. 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 by considering the chemical composition of the macromolecules of the cell. This implies that roughly half the dry mass comes from the carbon content of these cells, a figure that reveals 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

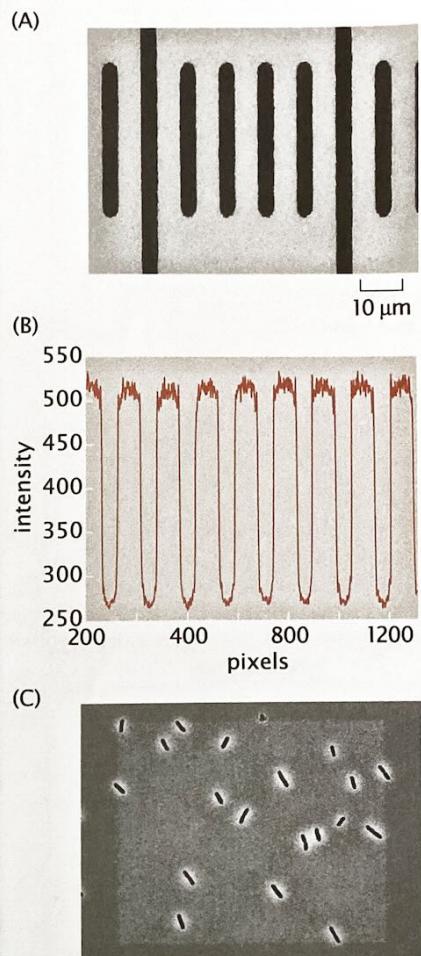


Figure 2.3: Sizing up *E. coli*.
 (A) Image of a graticule at 100 \times magnification. (B) Matlab plot of the intensity for a horizontal cut through the image. (C) Phase contrast image of a field of view with several *E. coli* cells taken at the same magnification as in (A).



ESTIMATE

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 $1 \text{ Da} \approx 1.6 \times 10^{-24} \text{ g}$, we have that our typical protein has a mass of $5 \times 10^{-20} \text{ g}$. The number of proteins per *E. coli* cell is estimated as

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

If we invoke the rough estimate that one-third of the proteins encoded 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 one-third by mass protein and two-thirds by mass RNA, facts that can be directly confirmed by the reader by inspecting the structural biology of ribosomes. As a result, we have

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

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 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

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

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 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

$$N_{\text{H}_2\text{O}} \approx \frac{0.7 \times 10^{-12} \text{ g}}{18 \text{ g/mol}} \times 6 \times 10^{23} \text{ molecules/mol}$$

$$\approx 2 \times 10^{10} \text{ water molecules.} \quad (2.4)$$

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

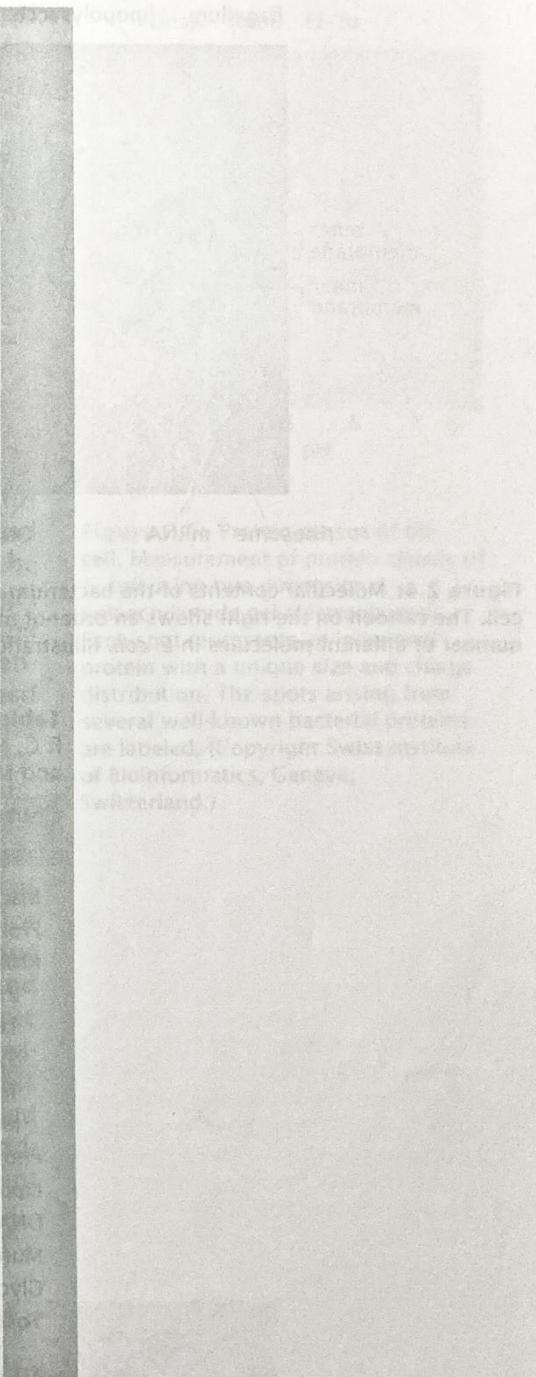
$$N_{\text{ions}} \approx \frac{(100 \times 10^{-3} \text{ mol}) \times (6 \times 10^{23} \text{ molecules/mol})}{10^{15} \mu\text{m}^3} \times 1 \mu\text{m}^3$$

$$= 6 \times 10^7. \quad (2.5)$$

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 2.4. 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.

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 2.5. 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 distributed through a tube-shaped polyacrylamide gel that has been polymerized to contain a stable pH gradient, and then an electric field is applied across the gel. The net charge on each protein depends on the pH and on the number and type of charged (protonatable) amino acid side chains that it contains. For example, the carboxylic acid group on aspartate will be negatively charged at high pH, but



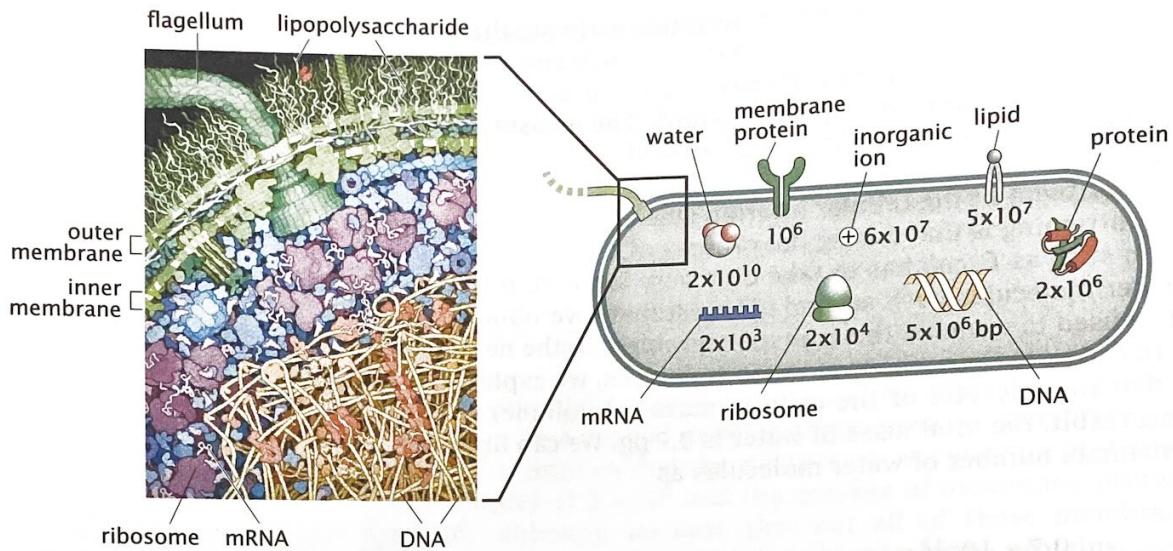


Figure 2.4: Molecular contents of the bacterium *E. coli*. The illustration on the left shows the crowded cytoplasm of the bacterial cell. The cartoon on the right shows an order-of-magnitude molecular census of the *E. coli* bacterium with the approximate number of different molecules in *E. coli*. (Illustration of the cellular interior courtesy of D. Goodsell.)

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

Substance	% of total dry weight	Number of molecules
Macromolecules		
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
Lipopopolysaccharide (outer membrane)	3.4	1.2×10^6
DNA	3.1	2
Murein (cell wall)	2.5	1
Glycogen (sugar storage)	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	

will pick up a hydrogen ion and will be neutral at low pH. Conversely, the amine group on lysine will be neutral at high pH, but will pick up a hydrogen ion and will be positively charged at low pH. The pH where a protein's charge is net neutral is called its "isoelectric point."

When a protein finds itself in a region of the gel where the pH is below its isoelectric point, there will be an excess of positive charge associated with the protein, and it will move toward the cathode when the electric field is applied. When a protein finds itself in a region of the gel where the pH is above its isoelectric point, it will have a net negative charge, and will therefore move toward the anode. When

all the proteins in the mixture have moved to the location of the pH where each is neutral, then all movement stops. At this point, "isoelectric focusing" is complete, and all of the proteins are arrayed along the tube-shaped gel in positions according 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. The detergent-soaked isoelectric focusing gel tube is placed at one end of a flat, square gel that also contains detergent, and an electric field is applied in a direction perpendicular to the first field. 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 the procedure described above, 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 2.5. 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.

More recently, several other methods have been brought to bear on the molecular census of cells. As shown in Figure 2.6, two of these methods are the use of mass spectrometry and fluorescence microscopy. In mass spectrometry, fragments of the many proteins contained within the cell are run through the mass spectrometer and their absolute abundances determined. The results of this technique applied to *E. coli* are shown in Figure 2.6(A). Alternatively, in fluorescence microscopy, a library of cells is created where each strain in this library expresses a particular protein from the *E. coli* proteome fused to a fluorescent protein. By calibrating the fluorescence corresponding to an individual fluorescent protein, one can measure the amount

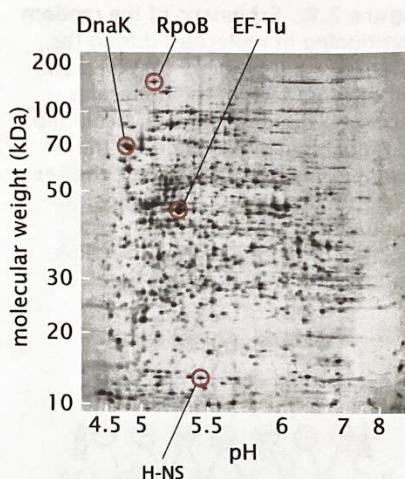


Figure 2.5: Protein census of the cell. Measurement of protein census of *E. coli* using two-dimensional polyacrylamide gel electrophoresis. Each spot represents an individual protein with a unique size and charge distribution. The spots arising from several well-known bacterial proteins are labeled. (Copyright Swiss Institute of Bioinformatics, Geneva, Switzerland.)

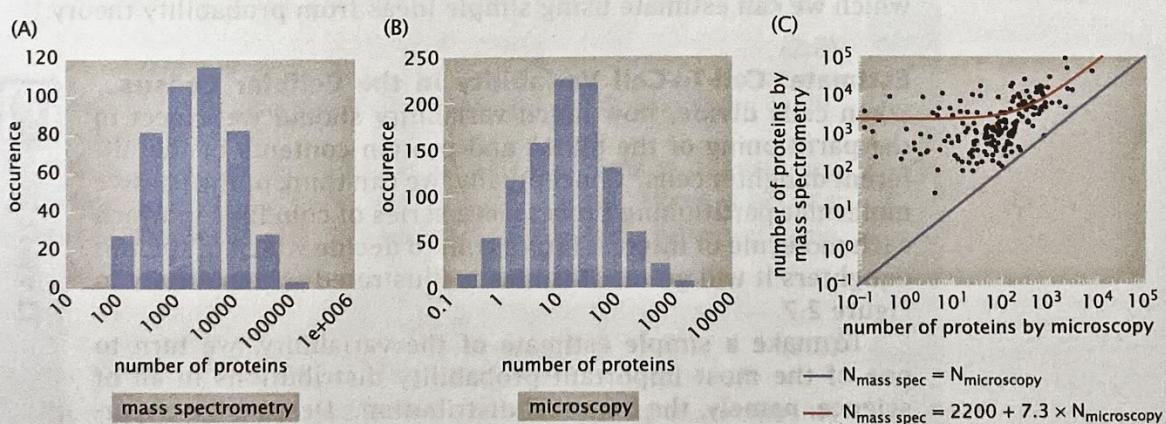
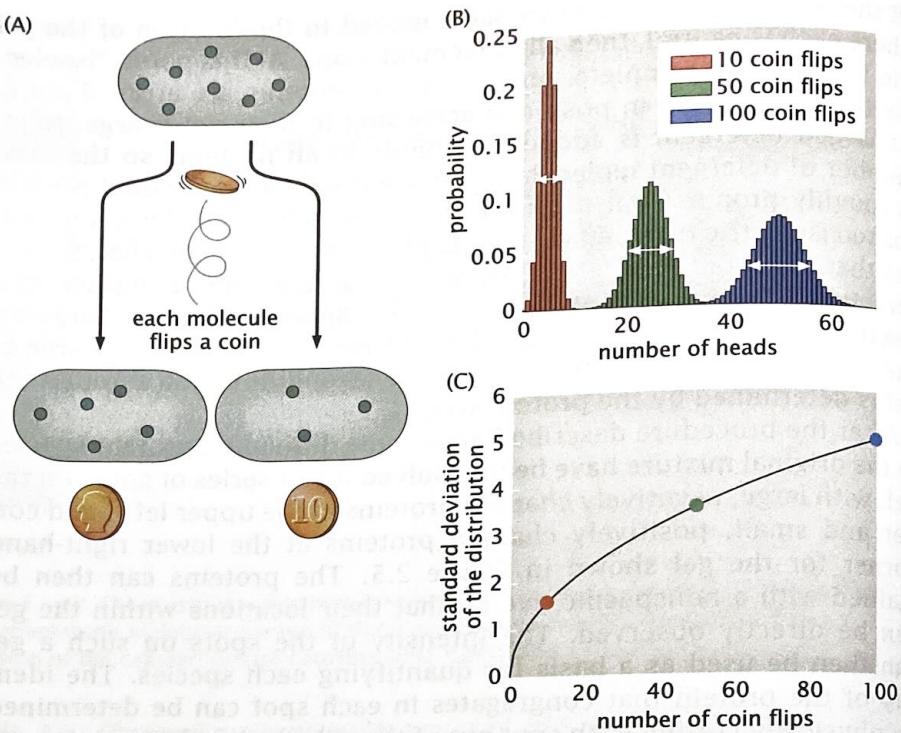


Figure 2.6: Protein census of *E. coli* using several techniques. A histogram of protein number in *E. coli* is shown from results using (A) mass spectrometry and (B) fluorescence microscopy with protein fusions to fluorescence proteins. (C) Comparison of mass spectrometry and fluorescence methods for the same collection of proteins showing a discrepancy between the two techniques as revealed by the fact that the best fit is not given by a line of slope one. (A, adapted from P. Lu et al., *Nat. Biotechnol.* 25:117, 2007; B, adapted from Y. Taniguchi et al., *Science* 329:533, 2010.)

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



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