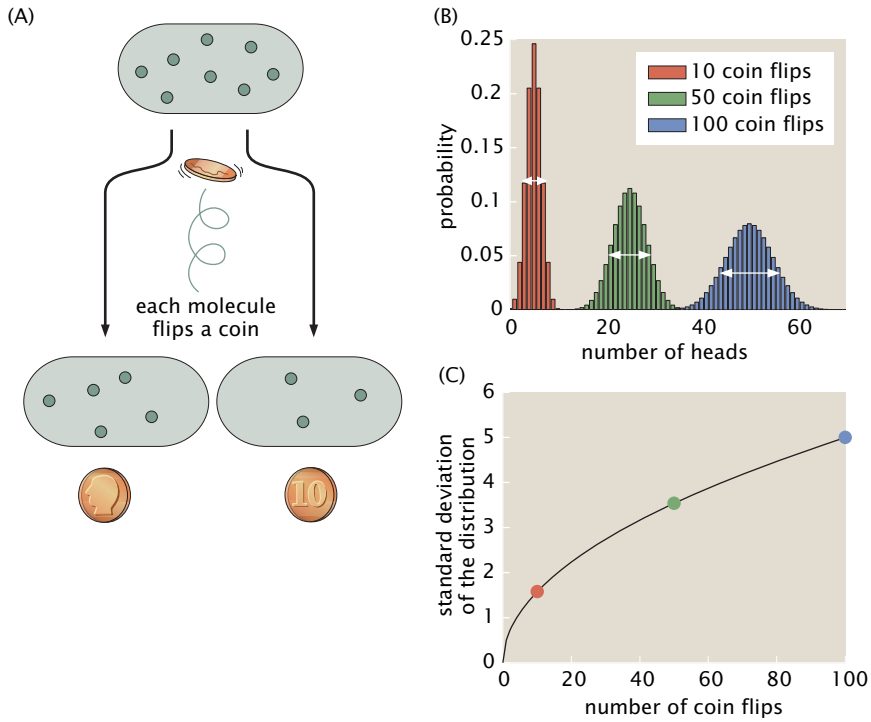

Figure 2.7: Schematic of the random partitioning of molecules during the process of cell division. (A) When the cell divides, each of the molecules chooses a daughter cell via a coin flip. (B) Probability distribution for the number of heads for different choices of the total number of coin flips. (C)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

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

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

ESTIMATE

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

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

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

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

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

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

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

$$\langle n_1 \rangle = Np. \tag{2.7}$$

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

$$\langle n_1^2 \rangle - \langle n_1 \rangle^2 = Npq, \tag{2.8}$$

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

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

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

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

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

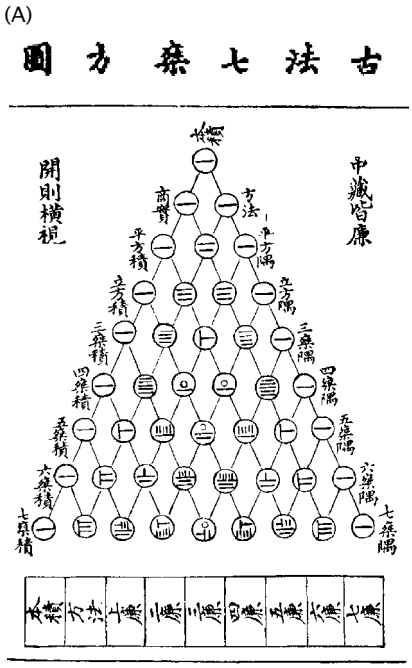

Figure 2.8: Binomial coefficients. (A) Though commonly known as Pascal's triangle, Chinese mathematicians had also earlier appreciated the symmetry of this collection of numbers. (B) The binomial coefficients that arise when raising $p + q$ to the power n, where n characterizes which row we are talking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

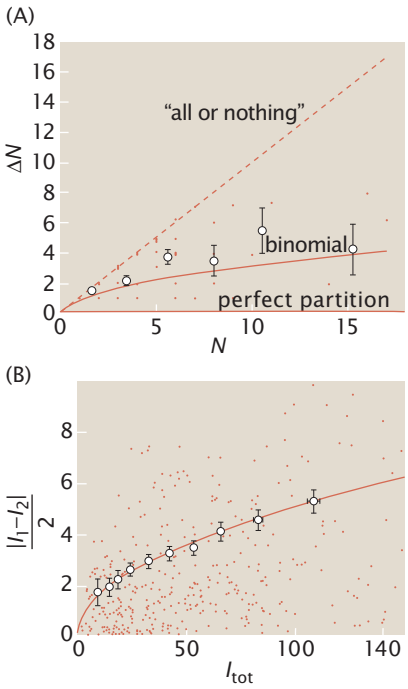

Figure 2.9: Binomial partitioning of mRNA and proteins in *E. coli* during cell division. (A) Difference ΔN in the number of mRNAs between the two daughters given that the mother has N mRNAs. The curves show three possible partitioning mechanisms involving "all or nothing" in which one daughter takes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

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

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

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

$$I_{\text{tot}} = \alpha N_{\text{tot}}, \tag{2.10}$$

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

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

COMPUTATIONAL EXPLORATION

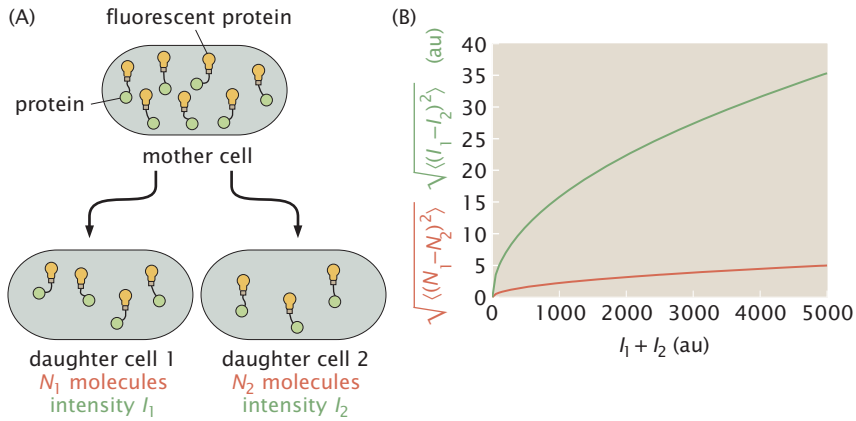

Figure 2.10: Intensity of the daughter cells as a result of binomial partitioning. (A) The mother cell starts out with a certain number of fluorescent proteins (represented by light bulbs) and upon division, these proteins are partitioned according to the binomial distribution. The number of fluorescent proteins and their corresponding total fluorescent intensity are related by the calibration factor α. (B) Plot of the intensity difference between the two daughters as a function of the intensity of the mother cell and the corresponding difference in partitioning of fluorescent molecules as a function of the total number of molecules in the mother cell.

actually determine N_{tot} by determining the unknown calibration factor. The idea of the experiment is to exploit the natural fluctuations resulting from imperfect partitioning of the fluorescent proteins between the two daughter cells. The key point is that for the case in which the partitioning between the two daughters is random (i.e., analogous to getting heads or tails in a coin flip), the size of those fluctuations depends in turn upon the total number of proteins being segregated between the two daughters. The larger the number of proteins, the larger the fluctuations. In particular, as shown in the problems at the end of the chapter, we can derive the simple and elegant result that the average difference in intensity between the two daughter cells is given by

$$\langle (I_1 - I_2)^2 \rangle = \alpha I_{tot}, \tag{2.11}$$

where I_1 and I_2 are the intensities of daughters 1 and 2, respectively, and I_{tot} is the total fluorescence intensity of the mother cell ($I_{tot} = I_1 + I_2$).

The simple analytic result of Equation 2.11 can perhaps be best appreciated through an appeal to a stochastic simulation. The idea is to "simulate" the dilution of fluorescent proteins during repeated cell divisions where all the mother cells are presumed to start with N copies of the fluorescent protein and the partitioning of those proteins to the two daughters is determined strictly by a simple binary random process (i.e., heads or tails of a coin flip of an honest coin). Conceptually, the algorithm for performing this simulation can be summarized as follows:

1. Choose the number of fluorescent molecules N in the mother cell and compute the intensity $I_{tot} = \alpha N$.
2. Generate N random numbers between 0 and 1. If a given random number is between 0 and 0.5, assign the protein to daughter 1 and otherwise assign the protein to daughter 2.
3. Compute the "intensity" of the daughters by multiplying N_1 and N_2 by α.
4. Perform the same operation for 100 cells (for example).
5. Choose a new N and repeat the process.

By considering the same division process again and again, for different choices of N, we can generate a curve that captures

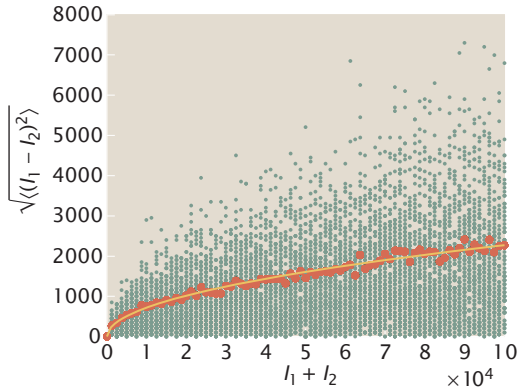

Figure 2.11: Simulation results of fluorescent protein dilution and counting experiment. For a given total number of proteins, N, in the mother cell, we simulate 100 different outcomes of partitioning by cellular division and plot their corresponding partitioning error (green dots). For each value of N, we can calculate the root mean square in partitioning (red dots) and fit them to the expected functional form given by the square root of Equation 2.11.

the statistics set forth above. In particular, for each and every N, 100 separate "divisions" were simulated and the results of each such division process are shown as the data points. For a given value of N, we can calculate the root mean square error in partitioning and its corresponding intensity, given by $\sqrt{\langle (I_1 - I_2)^2 \rangle}$. To see an example of the outcome of such a simulation, examine Figure 2.11.

Examples of Matlab code that could be used to perform this 'Computational Exploration' can be found on the book's, website.

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 2.4. This figure gives a view 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 2.4 can be further appreciated by converting these numbers into concentrations. To do so, we recall that the volume of an *E. coli* cell is roughly 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 2.12. 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

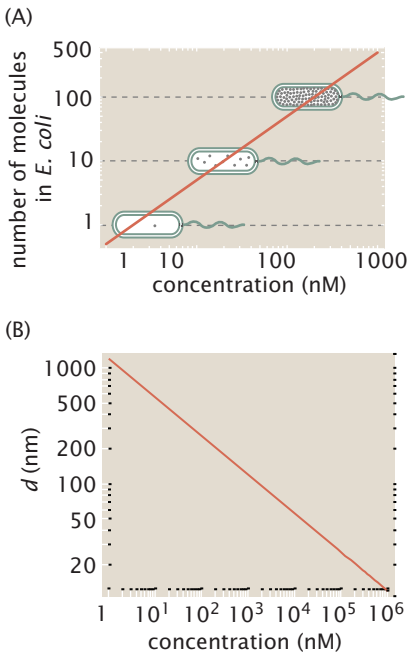

Figure 2.12: Physical interpretation of concentration. (A) Concentration in *E. coli* units: number of copies of a given molecule in a volume the size of an *E. coli* cell as a function of the concentration. (B) Concentration expressed in units of typical distance d between neighboring molecules measured in nanometers.