1 Proteins Don't Know Biology

1.1 PROLOGUE: STATISTICAL PHYSICS OF CANDY, DIRT, AND BIOLOGY

By the time you finish this book, hopefully you will look at the world around you in a new way. Beyond biomolecules, you will see that statistical phenomena are at work almost everywhere. Plus, you will be able to wield some impressive jargon and equations.

1.1.1 Candy

Have you ever eaten trail mix? A classic variety is simply a mix of peanuts, dried fruit, and chocolate candies. If you eat the stuff, you'll notice that you usually get a bit of each ingredient in every handful. That is, unsurprisingly, trail mix tends to be well mixed. No advanced technology is required to achieve this. All you have to do is shake.

To understand what's going on, let's follow a classic physics strategy. We'll simplify to the essence of the problem—the candy. I'm thinking of my favorite discoidally shaped chocolate candy, but you are free to imagine your own. To adopt another physics strategy, we'll perform a thought experiment. Imagine filling a clear plastic bag with two different colors of candies: first blue, then red, creating two layers. Then, we'll imagine holding the bag upright and shaking it (yes, we've sealed it) repeatedly. See Figure 1.1.

What happens? Clearly the two colors will mix, and after a short time, we'll have a fairly uniform mixture of red and blue candies.

If we continue shaking, not much happens—the well-mixed "state" is stable or "equilibrated." But how do the red candies know to move down and the blue to move up? And if the two colors are really moving in different directions, why don't they switch places after a long time?

Well, candy clearly doesn't think about what it is doing. The pieces can only move randomly in response to our shaking. Yet somehow, blind, random (nondirected) motion leads to a net flow of red candy in one direction and blue in the other. This is nothing other than the power of diffusion, which biomolecules also "use" to accomplish the needs of living cells. Biomolecules, such as proteins, are just as dumb as candy—yet they do what they need to do and get where they need to go. Candy mixing is just a simple example of a random process, which must be described statistically like many biomolecular processes.

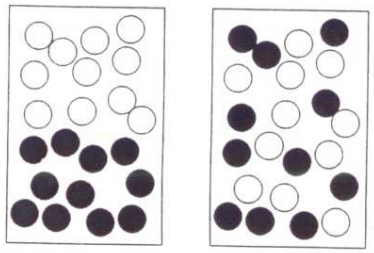

FIGURE 1.1 Diffusion at work. If a bag containing black and white candies is shaken, then the two colors get mixed, of course. But it is important to realize the candies don't know where to go in advance. They only move randomly. Further, once mixed, they are very unlikely to unmix spontaneously.

PROBLEM 1.1

Consider the same two-color candy experiment performed twice, each time with a different type of candy: once with smooth, unwrapped candy of two colors, and a second time using candy wrapped with wrinkly plastic. How will the results differ between wrapped and unwrapped candy? Hint: Consider what will happen based on a small number of shakes.

1.1.2 Clean Your House, Statistically

One of the great things about statistical physics is that it is already a part of your life. You just need to open your eyes to it.

Think about dirt. In particular, think about dirt on the floor of your house or apartment—the sandy kind that might stick to your socks a bit. If you put on clean socks and walk around a very dirty room, your socks will absorb some of that dirt—up to the point that they get "saturated." (Not a pleasant image, but conceptually useful!) With every additional step in the dirty room, some dirt may come on to your socks, but an approximately equal amount will come off. This is a kind of equilibrium.

Now walk into the hallway, which has just been swept by your hyper-neat housemate. Your filthy socks are now going to make that hallway dirty—a little. Of course, if you're rude enough to walk back and forth many times between clean and dirty areas, you will help that dirt steadily "diffuse" around the house. You can accelerate this process by hosting a party, and all your friends will transport dirt all over the house (Figure 1.2).

On the other hand, your clean housemate might use the party to his advantage. Assuming no dirt is brought into the house during the party (questionable, of course, but pedagogically useful), your housemate can clean the house without leaving his room! His strategy will be simple—to constantly sweep his own room, while allowing people in and out all the time. People will tend to bring dirt in to the clean room from the rest of the house, but they will leave with cleaner feet, having shed some dirt and picking up little in return. As the party goes on, more and more dirt will come

FONTE: Statistical Physics of Biomolecules: An Introduction; D. M. Zuckerman

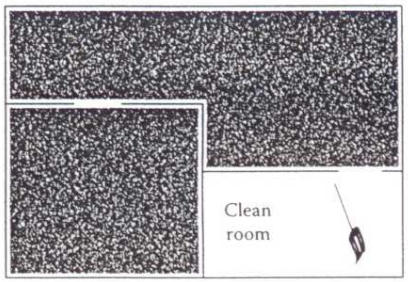

FIGURE 1.2 Clean your house statistically. A sweeper can stay in his room and let the dirt come to him, on the feet of his guests. If he keeps sweeping, dirt from the rest of the house can be removed. Similarly, proteins that only occupy a tiny fraction of the volume of a cell or beaker can "suck up" ligands that randomly reach them by diffusion.

into the clean room from people circulating all over the house, but there will be no compensating outflow of dirt. (Practical tip: very young guests can accelerate this process substantially, if they can be kept inside.)

PROBLEM 1.2

Consider a house with two rooms, one clean and one dirty. Imagine a number of people walk around randomly in this house, without bringing in more dirt. (a) Qualitatively, explain the factors that will govern the rate at which dirt travels from the clean to the dirty room—if no sweeping or removal of dirt occurs. (b) How does sweeping affect the process? (c) What if newly arriving guests bring in dirt from outside?

1.1.3 More Seriously...

In fact, the two preceding examples were very serious. If you understood them well, you can understand statistical biophysics. The candy and dirt examples illustrate fundamental issues of equilibrium, and the dynamical approach to equilibrium.

Everywhere in nature, dynamical processes occur constantly. In fact, it is fair to say that the execution of dynamics governed by forces is nature's only work. For instance, molecules move in space, they fluctuate in shape, they bind to one another, and they unbind. If a system is somewhat isolated, like the rooms of a closed house, an equilibrium can occur. The equilibrium can be changed by changing the "external conditions" like the total amount of dirt in the house or the size of the house. A biomolecular system—whether in a test tube or a living cell—can similarly be changed in quite a number of ways: by altering the overall concentration(s), changing the temperature or pH, or covalently changing one of the molecules in the system.

We won't be able to understand molecular systems completely, but hopefully we can understand the underlying statistical physics and chemistry. The principles described in this book are hardly limited to biology (think of all the candy and dirt in the world!), but the book is geared to an audience with that focus.

1.2 GUIDING PRINCIPLES

To get your mind wiggling and jiggling appropriately, let's discuss some key ideas that will recur throughout the book.

1.2.1 Proteins Don't Know Biology

Biology largely runs on the amazing tricks proteins can perform. But when it comes down to it, proteins are simply molecules obeying the laws of physics and chemistry. We can think of them as machines, but there's no ghost inside. Proteins are completely inanimate objects, whose primary role is to fluctuate in conformation (i.e., in shape or structure). Biology, via evolution, has indeed selected for highly useful structural fluctuations, such as binding, locomotion, and catalysis. However, to understand these highly evolved functions in a given molecule or set of molecules, it is very informative to consider their spontaneous "wigglings and jigglings," to paraphrase the physicist Richard Feynman. To put the idea a slightly different way: Biology at molecular lengthscales is chemistry and physics. Therefore, you can hope to understand the principles of molecular biophysics with a minimum of memorization and a maximum of clarity.

1.2.2 Nature Has Never Heard of Equilibrium

The fancy word "equilibrium" can mislead you into thinking that some part of biology, chemistry, or physics could be static. Far from it: essentially everything of scientific interest is constantly moving and fluctuating. Any equilibrium is only apparent, the result of a statistical balance of multiple motions (e.g., candy shaken upward vs. down). So an alternative formulation of this principle is "Nature can't do statistical calculations." Rather, there tend to be enough "realizations" of any process that we can average over opposing tendencies to simplify our understanding.

Like proteins, nature is dumb. Nature does not have access to calculators or computers, so it does not know about probabilities, averages, or standard deviations. Nature only understands forces: push on something and it moves. Thus, this book will frequently remind the reader that even when we are using the comfortable ideas of equilibrium, we are actually talking about a balance among dynamical behaviors. It is almost always highly fruitful to visualize the dynamic processes underlying any equilibrium.

1.2.2.1 Mechanistic Promiscuity?

While we're on the subject, it's fair to say that nature holds to no abstract theories at all. Nature is not prejudiced for or against particular "mechanisms" that may inspire controversy among humans. Nature will exploit—indeed, cannot help but exploit—any mechanism that passes the evolutionary test of continuing life. Therefore, this book attempts to steer clear of theorizing that is not grounded in principles of statistical physics.

1.2.3 Entropy Is Easy

Entropy may rank as the worst-explained important idea in physics, chemistry, and biology. This book will go beyond the usual explanation of entropy as uncertainty and beyond unhelpful equations, to get to the root meaning in simple terms. We'll also see what the unhelpful equations mean, but the focus will be on the simplest (and, in fact, most correct) explanation of entropy. Along the way, we'll learn that understanding "free" energy is equally easy.

1.2.4 Three Is the Magic Number for Visualizing Data

We can only visualize data concretely in one, two, or three dimensions. Yet large biomolecules "live" in what we call "configuration spaces," which are very high dimensional—thousands of dimensions, literally. This is because, for example, if a protein has 10,000 atoms, we need 30,000 numbers to describe a single configuration (the x, y, and z values of every atom). The net result is that even really clever people are left to study these thousands of dimensions in a very partial way, and we'll be no different. However, we do need to become experts in simplifying our descriptions, and in understanding what information we lose as we do so.

1.2.5 Experiments Cannot Be Separated from "Theory"

The principles we will cover are not just of interest to theorists or computationalists. Rather, because they are actually true, the principles necessarily underpin phenomena explored in experiments. An auxiliary aim of this book, then, is to enable you to better understand many experiments. This connection will be explicit throughout the book.

1.4 MOLECULAR PROLOGUE: A DAY IN THE LIFE OF BUTANE

Butane is an exemplary molecule, one that has something to teach all of us. In fact, if we can fully understand butane (and this will take most of the book!) we can understand most molecular behavior. How can this be? Although butane (*n*-butane to the chemists) is a very simple molecule and dominated by just one degree of freedom, it consists of 14 atoms (C_4H_{10}). Fourteen may not sound like a lot, but when you consider that it takes 42 numbers to describe a single butane configuration and its orientation in space (x, y, and z values for each atom), that starts to seem less simple. We can say that the "configuration space" of butane is 42-dimensional. Over time, a butane molecule's configuration can be described as a curve in this gigantic space.

Butane's configuration is most significantly affected by the value of the central torsion or dihedral angle, ϕ. (A dihedral angle depends on four atoms linked sequentially by covalent bonds and describes rotations about the central bond, as shown in Figure 1.3. More precisely, the dihedral is the angle between two planes—one formed by the first three atoms and the other by second three.) Figure 1.3 shows only the carbon atoms, which schematically represent the sequence of four chemical groups (CH_3,CH_2,CH_2,CH_3).

In panel (a) of Figure 1.3, we see a computer simulation trajectory for butane—specifically, a series of ϕ values at equally spaced increments in time. This is a day

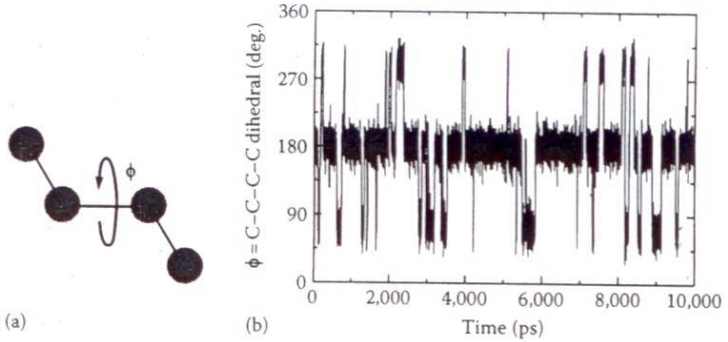

FIGURE 1.3 Butane and its life. Panel (a) shows a simplified representation of the butane molecule, with only the carbon atoms depicted. Rotations about the central C–C bond (i.e., changes in the ϕ or C–C–C–C dihedral angle) are the primary motions of the molecule. The "trajectory" in panel (b) shows how the value of the dihedral changes in time during a molecular dynamics computer simulation of butane.

in the life of butane. Well, it's really just a tiny fraction of a second, but it tells the whole story. We can see that butane has three states—three essentially discrete ranges of ϕ values—that it prefers. Further, it tends to stay in one state for a seemingly random amount of time, and then make a rapid jump to another state. In the following chapters, we will study all this in detail: (1) the reason for the quasi-discrete states; (2) the jumps between states; (3) the connection between the jumps and the observed equilibrium populations; and, further, (4) the origin of the noise or fluctuations in the trajectory.

What would we conclude if we could run our computer simulation—essentially a movie made from many frames or snapshots—for a very long time? We could then calculate the average time interval spent in any state before jumping to other states, which is equivalent to knowing the rates for such isomerizations (structure changes). We could also calculate the equilibrium or average fractional occupations of each state. Such fractional occupations are of great importance in biophysics.

A simple, physically based way to think about the trajectory is in terms of an energy profile or landscape. For butane, the landscape has three energy minima or basins that represent the states (see Figure 1.4). As the famous Boltzmann factor (e^{-E/k_BT}, detailed later) tells us, probability decreases exponentially with higher energy E. Thus, it is relatively rare to find a molecule at the high points of the energy landscape. When such points are reached, a transition to another state can easily occur, and thus such points are called barriers. Putting all this together, the trajectory tends to fluctuate around in a minimum and occasionally jump to another state. Then it does the same again.

Another technical point is the diversity of timescales that butane exhibits. For instance, the rapid small-scale fluctuations in the trajectory are much faster than the transitions between states. Thinking of butane's structure, there are different types of motions that will occur on different timescales: the fast bond-length and bond-angle fluctuations, as opposed to the slow dihedral-angle fluctuations.

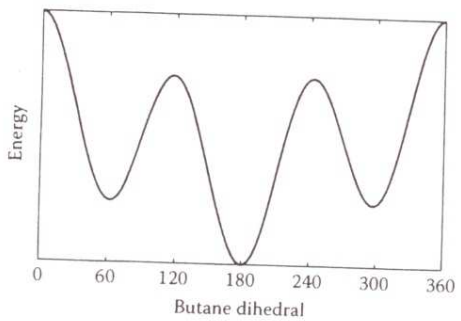

FIGURE 1.4 The "energy" landscape of butane. Butane spends most of its time in one of the three energy minima shown, as you can verify by examining the trajectory in Figure 1.3. Later, we will learn that a landscape like this actually represents a "free energy" because it reflects the relative population of each state—and even of each ϕ value.

1.4.1 Exemplary by Its Stupidity

Details aside, butane is exemplary for students of biophysics in the sense that it's just as dumb as a protein—or rather, proteins are no smarter than butane. All any molecule can do is jump between the states that are determined by its structure (and by interactions with nearby molecules). Protein—or RNA or DNA—structures happen to be much more complicated and tuned to permit configurational fluctuations with important consequences that permit us to live!

One of the most famous structural changes occurs in the protein hemoglobin, which alters its shape due to binding oxygen. This "allosteric" conformational change facilitates the transfer of oxygen from the lungs to the body's tissues. We will discuss principles of allosteric conformational changes in Chapter 9.

1.5 WHAT DOES EQUILIBRIUM MEAN TO A PROTEIN?

Proteins don't know biology, and they don't know equilibrium either. A protein "knows" its current state—its molecular configuration, atomic velocities, and the forces being exerted on it by its surroundings. But a protein will hardly be affected by the trillions of other proteins that may co-inhabit a test tube with it. So how can a scientist studying a test tube full of protein talk about equilibrium?

There are two important kinds of equilibrium for us. One is an inter molecular equilibrium that can occur between different molecules, which may bind and unbind from one another. Within any individual molecule, there is also an internal equilibrium among different possible configurations. Both types of equilibriums are worth previewing in detail.

1.5.1 Equilibrium among Molecules

Intermolecular equilibrium occurs among species that can bind with and unbind from one another. It also occurs with enzymes, proteins that catalyze chemical changes in their binding partners.

Imagine, as in Figure 1.5, a beaker filled with many identical copies of a protein and also with many ligand molecules that can bind the protein—and also unbind from it. Some fraction of the protein molecules will be bound to ligands, depending on the affinity or strength of the interaction. If this fraction does not change with time (as will usually be the case, after some initial transient period), we can say that it represents the equilibrium of the binding process. Indeed, the affinity is usually quantified in terms of an equilibrium constant, K_d, which describes the ratio of unbound-to-bound molecules and can be measured in experiments.

But how static is this equilibrium? Let's imagine watching a movie of an individual protein molecule. We would see it wiggling and jiggling around in its aqueous solution. At random intervals, a ligand would diffuse into view. Sometimes, the ligand might just diffuse away without binding, while at other times it will find the protein's binding site and form a complex. Once bound, a ligand will again unbind (after some random interval influenced by the binding affinity) and the process will repeat. So where is the equilibrium in our movie?

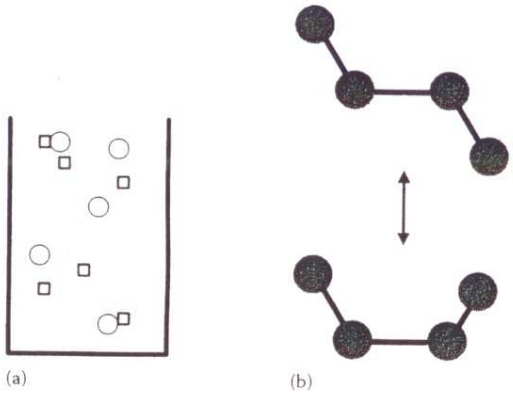

FIGURE 1.5 Two kinds of equilibrium. (a) A beaker contains two types of molecules that can bind to one another. In equilibrium, the rate of complex formation exactly matches that of unbinding. (b) A conformational equilibrium is shown between two (out of the three) states of butane. Transitions will occur back and forth between the configurations, as we have already seen in Figure 1.3.

The equilibrium only exists in a statistical sense, when we average over the behavior of many proteins and ligands. More fundamentally, when we consider all the molecules in the beaker, there will be a certain rate of binding, largely governed by the molecular concentrations—that is, how frequently the molecules diffuse near each other and also by the tendency to bind once protein and ligand are near. Balancing this will be the rate of unbinding (or the typical complex lifetime), which will be determined by the affinity—the strength of the molecular interactions. In equilibrium, the total number of complexes forming per second will equal the number unbinding per second.

Equilibrium, then, is a statistical balance of dynamical events. This is a key idea, a fundamental principle of molecular biophysics. We could write equations for it (and later we will), but the basic idea is given just as well in words.

1.5.2 Internal Equilibrium

The same ideas can be applied to molecules that can adopt more than one geometric configuration. Butane is one of the simplest examples, and the equilibrium between two of its configurational states is schematized in Figure 1.5. But all large biomolecules, like proteins, DNA, and RNA, can adopt many configurations—as can many much smaller biomolecules, such as drugs.

Again, we can imagine making a movie of a single protein, watching it interconvert among many possible configurations—just as we saw for butane in Figure 1.3. This is what proteins do: they change configuration in response to the forces exerted on them. To be a bit more quantitative, we could imagine categorizing all configurations in our movie as belonging to one of a finite number of states, $i = 1, 2, \ldots$. If our movie was long enough so that each state was visited many times, we could even reliably estimate the average fraction of time spent in each state, p_i (i.e., the fractional population). Since every configuration belongs to a unique state, all the fractions would sum to one, of course.

In a cell, proteins will typically interact with many other proteins and molecules that are large when compared to water. The cell is often called a crowded environment. Nevertheless, one can still imagine studying a protein's fluctuations, perhaps by making a movie of its motions, however complex they may be.

1.5.3 Time and Population Averages

Here's an interesting point we will discuss again, later on. Dynamical and equilibrium measurements must agree on populations. To take a slightly different perspective, we can again imagine a large number of identical proteins in a solution. We can also imagine taking a picture ("snapshot") of this entire solution. Our snapshot would show what is called an ensemble of protein configurations. As in the dynamical case, we could categorize all the configurations in terms of fractional populations, now denoted as $\hat{p}_i$.

The two population estimates must agree, that is, we need to have $\hat{p}_i = p_i$. Proteins are dumb in the sense that they can only do what their chemical makeup and configuration allow them to do. Thus, a long movie of any individual protein will have identical statistical properties to a movie of any other chemically identical copy of that protein. A snapshot of a set of proteins will catch each one at a random point in its own movie, forcing the time and ensemble averages to agree.

PROBLEM 1.3

Apply this same argument to the ligand-binding case, showing a connection between a movie and the equilibrium fraction of bound proteins.

1.5.3.1 A Dynamical Description Has More Information Than an Equilibrium Picture

This is a simple but important point. Although dynamical measurements (i.e., from movies) must agree with equilibrium/ensemble measurements, ensemble measurements lack dynamical information. In particular, we cannot learn the rates of binding or of interconversion among configurations based on equilibrium measurements. To give a simple example, an equilibrium description of one person's sleep habits might indicate she sleeps 7/24 of the day on average. But we won't know whether that means 7 h on the trot, or 5 h at night and a 2 h afternoon nap.

1.6 A WORD ON EXPERIMENTS

Experiments can be performed under a variety of conditions that we can understand based on the preceding discussion. Perhaps the most basic class of experiments is the ensemble equilibrium measurement (as opposed to a single-molecule measurement or a nonequilibrium measurement). Examples of ensemble equilibrium measurements are typical NMR (nuclear magnetic resonance) and x-ray structure determination

experiments. These are ensemble measurements since many molecules are involved. They are equilibrium measurements since, typically, the molecules have had a long time to respond to their external conditions. Of course, the conditions of an NMR experiment—aqueous solution—are quite different from x-ray crystallography where the protein is in crystalline form. Not surprisingly, larger motions and fluctuations are expected in solution, but some fluctuations are expected whenever the temperature is nonzero, even in a crystal. (Interestingly, protein crystals tend to be fairly wet: they contain a substantial fraction of water, so considerable fluctuations can be possible.) All this is not to say that scientists fully account for these fluctuations when they analyze data and publish protein structures, either based on x-ray or NMR measurements, but you should be aware that these motions must be reflected in the raw (pre-analysis) data of the experiments.

Nonequilibrium measurements refer to studies in which a system is suddenly perturbed, perhaps by a sudden light source, temperature jump, or addition of some chemical agent. As discussed above, if a large ensemble of molecules is in equilibrium, we can measure its average properties at any instant of time and always find the same situation. By a contrasting definition, in nonequilibrium conditions, such averages will change over time. Nonequilibrium experiments are thus the basis for measuring kinetic processes—that is, rates.

Recent technological advances now permit single-molecule experiments of various kinds. These measurements provide information intrinsically unavailable when ensembles of molecules are present. First imagine two probes connected by a tether consisting of many molecules. By tracking the positions of the probes, one can only measure the average force exerted by each molecule. But if only a single molecule were tethering the probes, one could see the full range of forces exerted (even over time) by that individual. Furthermore, single-molecule measurements can emulate ensemble measurements via repeated measurements.

Another interesting aspect of experimental measurements is the implicit time averaging that occurs; that is, physical instruments can only make measurements over finite time intervals and this necessarily involves averaging over any effects occurring during a brief window of time. Think of the shutter speed in a camera: if something is moving fast, it appears as a blur, which reflects the time-averaged image at each pixel.

To sum up, there are two key points about experiments. First, our "theoretical" principles are not just abstract but are 100% pertinent to biophysical behavior measured in experiments. Second, the principles are at the heart of interpreting measured data. Not surprisingly, these principles also apply to the analysis of computer simulation data.

1.7 MAKING MOVIES: BASIC MOLECULAR DYNAMICS SIMULATION

The idea of watching a movie of a molecular system is so fundamental to the purpose of this book that it is worthwhile to sketch the process of creating such a movie using computer simulation. While there are entire books devoted to molecular simulation of various kinds, we will present a largely qualitative picture of molecular dynamics (MD) simulation. Other basic ideas regarding molecular simulation will be described briefly in Chapter 12.

MD simulation is the most literal-minded of simulation techniques, and that is partly why it is the most popular. Quite simply, MD employs Newton's second law ($f = ma$) to follow the motion of every atom in a simulation system. Such a system typically includes solute molecules such as a protein and possibly a ligand of that protein, along with solvent like water and salt. (There are many possibilities, of course, and systems like lipid membranes are regularly simulated by MD.) In every case, the system is composed of atoms that feel forces. In MD, these forces are described classically—that is, without quantum mechanics, so the Schrödinger equation is not involved. Rather, there is a classical "forcefield" (potential energy function) U, which is a function of all coordinates. The force on any atomic coordinate, x_i, y_i, or z_i for atom i, is given by a partial derivative of the potential: for instance, the y component of force on atom i is given by $-\partial U/\partial y_i$.

MD simulation reenacts the simple life of atom: an atom will move at its current speed unless it experiences a force that will accelerate or decelerate it. To see how this works (and avoid annoying subscripts), we'll focus on the single coordinate x, with velocity $v = dx/dt$ and force $f = -dU/dx$. We can use Newton's law ($a = dv/dt = f/m$) to describe the approximate change in speed a short time Δt after the current time t. If we write $dv/dt \simeq \Delta v/\Delta t$, we find that $a\Delta t = v(t + \Delta t) - v(t)$. We can then write an equation for the way velocity changes with time due to a force:

$$v(t + \Delta t) = v(t) + a(t)\Delta t = v(t) + \left[\frac{f(t)}{m}\right]\Delta t. \tag{1.1}$$

As in freshman physics, we can also integrate this equation under the assumption that speed and acceleration are constant over the small time interval Δt. The integration yields an equation for the position at time $t + \Delta t$, and should also look familiar:

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \left[\frac{f(t)}{2m}\right](\Delta t)^2, \tag{1.2}$$

where $f(t)/m = a(t)$.

It is straightforward to read such an equation: the x coordinate at time $t + \Delta t$ is the old coordinate plus the change due to velocity, as well as any additional change if that velocity is changing due to a force.

To implement MD on a computer, one needs to calculate forces and to keep track of positions and velocities. When this is done for all atoms, a trajectory (like that shown in Figure 1.3 for butane) can be created. More schematically, take a look at Figure 1.6. To generate a single new trajectory/movie-frame for butane without any solvent, one needs to repeat the calculation of Equation 1.2 42 times—once for each of the x, y, and z components of all 14 atoms! Sounds tedious (it is), but this is a perfect task for a computer. Any time a movie or trajectory of a molecule is discussed, you should imagine this process.

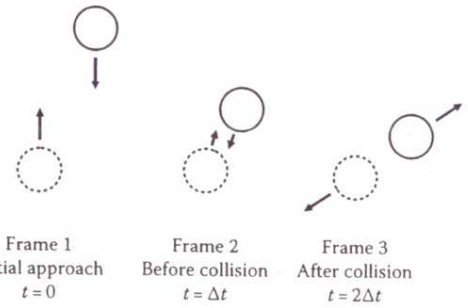

FIGURE 1.6 Making movies by molecular dynamics. Two colliding atoms are shown, along with arrows representing their velocities. The atoms exert repulsive forces on one another at close range, but otherwise are attractive. Note that the velocities as shown are not sufficient to determine future positions of the atoms. The forces are also required.

PROBLEM 1.4

Make your own sketch or copy of Figure 1.6. Given the positions and velocities as shown, sketch the directions of the forces acting on each atom at times $t = 0$ and $t = \Delta t$. Can you also figure out the forces at time $t = 2\Delta t$?