Statistical Mechanics Primer

David Van Valen

January 7, 2007

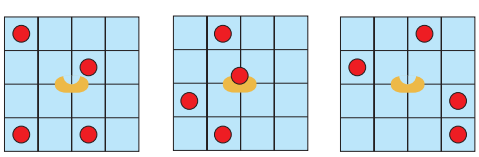
As the data produced by experimental biologists becomes more quantitative, there becomes a need for more quantitative models. There are many ways to build models (rate equations and Markov processes come to mind), and each method has time and place when they should be used. In this set of notes, we explore statistical mechanics as a method to build models for biology. Gene regulation, receptor-ligand binding, and macromolecular crowding are just some of the cases that can be treated using statistical mechanics. Given its utility in biology and its widespread use in this course, it is important to have a good handle on statistical mechanics. Most of what you will need to know is outlined in this set of notes. These notes are divided as follows. First, we first introduce concepts such as microscopic states and entropy. We then derive the Boltzmann distribution, the heart and soul of statistical mechanics. Finally, we present several applications of statistical mechanics to different problems in biology to give a flavor for how these kinds of models are used.

1 Microstates and Entropy

Bulk properties of systems are of great interest when performing experiments in physics. When dealing with gases, it is more useful to know the volume, pressure, and temperature than the coordinates and velocities of the individual gas molecules. This is also the case in biology. The average expression of a gene in a population of cells is a quantity that is often measured. Naturally, the macroscopic properties are influenced by what is happening at the microscopic level. The fast molecules in the gas help determine a gas' temperature just as the position and binding partners of transcription factors determine the level of gene expression. Here is where statistical mechanics comes in. Statistical mechanics helps connect microscopic behavior with macroscopic phenomena. Before proceeding, we need to understand the concept of a "microstate."

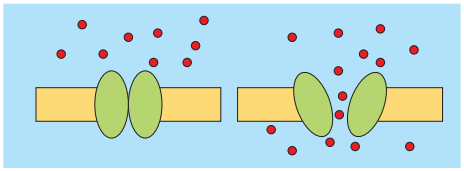
Microstates are similar to the matrix. No one can tell you what a microstate is; you have to see it for yourself. Wikipedia defines a microstate as "a specific detailed microscopic configuration of a system that [... a] system visits in the course of its thermal fluctuations." While this definition is correct, it is more illuminating to look at several concrete cases.

Figure 1 shows three cases where we can define microstates. The first figure shows a lattice model of a receptor in a solution of ligands. In this case, each spatial configuration of the ligands represents a different microstate. Some microstates, like the ones with ligand bound to the receptor, have more functional significance for others. For fun, let us figure out how many microstates there are for this figure. There are 16 sites on the lattice and 1 site for the receptor - this gives 17 total sites for the 4 ligands. There are 17 ways to pick the first ligand, 16 ways to pick the second, 15 ways to pick the third, and 14 ways to pick the fourth site. This gives $17 * 16 * 15 * 14 = 57120$. There is an error with this calculation - we forgot to account for the fact that the ligands are indistinguishable. The method of counting treated the ligands as if they were labeled 1, 2, 3, and 4, but in reality we can't distinguish one ligand from another. There are $4 * 3 * 2 * 1 = 4! = 24$ ways to label the 4 ligands, so we must divide our answer by 24. This gives 2380 as the number of possible microstates. As a general rule, the number of ways to choose R identical objects from a set of N objects is given by $\binom{N}{R} = \frac{N!}{R!(N-R)!}$.



(a) Molecules in solution.

(b) Protein folding.



(c) Ion Channels.

Figure 1: Different examples of microstates.

The second figure shows a schematic of a protein. The microstates in this case are the conformations of the backbone. If the properly folded protein performs some function, then we have a clear dependence of a bulk quantity (the activity of the protein) on which microstates are occupied. It is interesting to note that with a continuous representation of the backbone, there are an uncountably infinite number of configurations for the backbone. We can count the number of microstates though, either by making a discrete model of the backbone or by making some changes to how we count. This nuance is not that important. The third figure shows an ion channel. There are only two microstates this system occupies - open and closed. A bulk quantity of interest in this case could be the probability a channel is open. This would determine the flux of ions into a cell, a process that has a wide range of effects from controlling volume to transmitting signals in the nervous system.

While the calculation above may seem elementary, it is actually of utmost importance to statistical mechanics. The number of available microstates will impact the equilibrium behavior of any system. It stands to reason that having some measure of how many microstates are in a system would be very useful. There is a problem with simply counting microstates - the result you get is not additive. Consider the case of ligands in solution - in that example there are 2380 possible microstates. If we expand our system to include an identical solution that is in a separate beaker, then the number of microstates become $2380^2 = 5664400$. It would be nice to have some measure that doubles when we double our system. We can get such a measure by taking the logarithm of the number of microstates. In other words, if Ω is the number of microstates, then we can define

$$S = k_B \log \Omega, \tag{1}$$

where k_B is Boltzmann's constant, to get an additive measure of the number of microstates. S is called the entropy. In our thought experiment, the entropy of the solution is $S = k_B \log 2380 = 7.77 k_B$ and when we double the system the entropy becomes $S = k_B \log 2380^2 = 2k_B \log 2380 = 15.54$. It can be shown that, up to a constant, this is the only definition that is additive.

2 The Boltzmann Distribution

Let us consider the following problem: suppose we have a small box in equilibrium with a large heat bath. What is the probability the box occupies a microstate with energy E? Let's start by stating what we know. Because the heat bath and system are in thermal equilibrium, they have the same temperature T. Let's

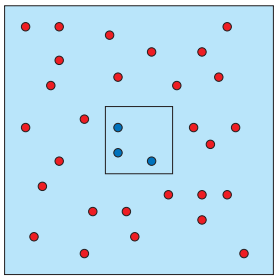


Figure 2: A small system in equilibrium with a heat bath.

denote the energy of the system as E and the energy of the heat bath as E_B. The total energy is given by

$$E_T = E_B + E. \tag{2}$$

The heat bath and system can exchange energy, but the total energy E_T must remain a constant. Because the heat bath is much larger than the system by assumption, then E should be much lower than E_B. We have already specified which microstate the box occupies. We have no such specification for the heat bath; all we know is that it has energy $E_B = E_T - E$. It could be in a multitude of microstates. The number of microstates accessible to it depends on its energy. Let $\Omega(E_B)$ denote the number of microstates of the heat bath when it has energy E_B.

What distinguishes one microstate of the heat bath from another? Nothing. It only makes sense that it is equally likely for the heat bath to be in one microstate as opposed to another. This concept is called *equal a priori probability* and is at the heart of statistical mechanics. Although the microstates are equally likely, the number of microstates differs depending on how much energy the heat bath as donated to the box. The probability of the box having energy E should be proportional to the number of microstates the heat bath has when it has donated E to the box. In other words,

$$p \propto \Omega(E_T - E). \tag{3}$$

As stated above E is small compared to E_T. This invites a taylor expansion, which we do after taking the logarithm of both sides. This gives

$$\begin{aligned}
\log p &\propto \log \Omega(E_T - E), && (4)\\
&\propto \log \Omega(E_T) + \left(\frac{\partial \log \Omega(x)}{\partial x}\right)_{x=E_T} (E_T - E), && (5)\\
&\propto \log \Omega(E_T) + \frac{1}{k_B} \underbrace{\left(\frac{\partial S(x)}{\partial x}\right)_{x=E_T}}_{T} (E_T - E), && (6)\\
&\propto \text{const} - \frac{E}{k_B T}. && (7)
\end{aligned}$$

This means

$$\begin{aligned}
p &\propto e^{-\frac{E}{k_B T}}, && (8)\\
&= \frac{1}{Z} e^{-\frac{E}{k_B T}}, && (9)
\end{aligned}$$

where

$$Z = \sum_{\text{states}} e^{-\frac{E}{k_B T}} \tag{10}$$

is a normalization constant called the partition function. For simplicity, people often let $\beta = \frac{1}{k_B T}$. $e^{-\beta E}$ is often called the Boltzmann factor; when one sums over microstates that have the same Boltzmann factor the result is often called the statistical weight. *This is the most important formula in statistical mechanics. If you remember anything from these notes, remember this.* This rule allows us to calculate the probability of the box being in any microstate. The box can represent anything - DNA and transcription factors, the conformation of a membrane, you name it. If you can calculate the statistical weight, you can use the toolkit of statistical mechanics.

Before proceeding, there is one major requirement that must be met before using this formula - the system under study must be in equilibrium. This is often not the case in biology. However, even if we limit ourselves to cases where we are justified in assuming equilibrium exists, statistical mechanics still very useful.

3 Applications

All of statistical mechanics can be summarized in $e^{-\beta\epsilon}$. Now that we have the rule for determining the probability of microstates, let us work out some biologically relevant examples.

3.1 Receptor-Ligand Binding

Lets consider the previous example of a lattice model of ligands binding to a receptor. Suppose there is one receptor, N ligands, and V sites in the solution. Binding to the receptor confers an energy bonus ϵ. Let's

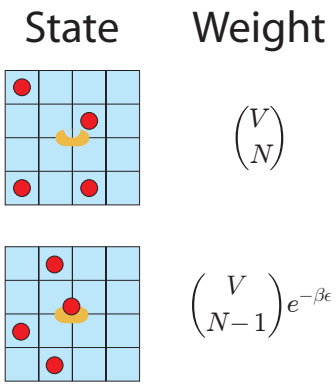


Figure 3: States and weights of ligands binding to a receptor

calculate the probability a ligand is bound to the receptor. This probability is given by

$$p_{bound} = \frac{\sum_{\text{bound states}} e^{-\beta E_{\text{state}}}}{\sum_{\text{bound states}} e^{-\beta E_{\text{state}}} + \sum_{\text{unbound states}} e^{-\beta E_{\text{state}}}} \tag{11}$$

Let us find the partition function by considering the unbound and bound states separately. If no ligand is bound to the receptor, then the energy is 0. This gives a Boltzmann factor of $e^0 = 1$. When we sum over states, the sum just becomes the number of microstates. How many microstates correspond to the unbound macrostate? There are V sites and we must choose N so there are $\binom{V}{N}$ microstates. The statistical weight is then $\binom{V}{N}$. Now suppose a ligand is bound to the receptor. The Boltzmann factor is $e^{-\beta\epsilon}$ and the sum over

states is $e^{-\beta\epsilon} \times \#$ microstates. There are now only $N-1$ ligands in solution, so the number of microstates is $\binom{V}{N-1}$. The statistical weight is then $\binom{V}{N-1}e^{-\beta\epsilon}$. Plugging in our expressions gives

$$\begin{aligned} p_{bound} &= \frac{\binom{V}{N-1}e^{-\beta\epsilon}}{\binom{V}{N-1}e^{-\beta\epsilon} + \binom{V}{N}}, \\ &= \frac{\frac{V!}{(N-1)!(V-N+1)!}e^{-\beta\epsilon}}{\frac{V!}{(N-1)!(V-N+1)!}e^{-\beta\epsilon} + \frac{V!}{N!(V-N)!}}, \\ &= \frac{1}{1 + \frac{V!(N-1)!(V-N+1)!}{N!(V-N)!V!}e^{\beta\epsilon}}, \\ &= \frac{1}{1 + \frac{V-N+1}{N}e^{\beta\epsilon}}. \end{aligned}$$

Let's work with the assumption that the solution is dilute. This means that $V >> N-1$. We can then rewrite the equation as

$$p_{bound} = \frac{1}{1 + \frac{V}{N}e^{\beta\epsilon}}, \tag{12}$$

$$= \frac{\frac{N}{V}}{e^{\beta\epsilon} + \frac{N}{V}}. \tag{13}$$

$$\tag{14}$$

Let's examine this equation to gain some intuition. First, note that $0 \leq \frac{N}{V} \leq 1$. If the binding is very favorable, then $\epsilon << 0$, $e^{\beta\epsilon} \approx 0$, and $p_{bound} \approx 1$. If the binding is very unfavorable, then $\epsilon >> 0$, $e^{\beta\epsilon} >> 1$, and $p_{bound} \approx 0$. All of this agrees with our intuition. It is important to note is that when ϵ is near k_BT,

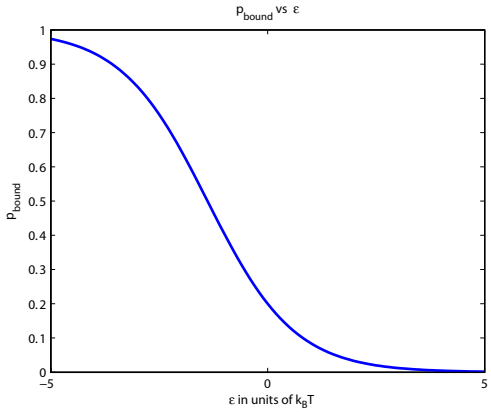


Figure 4: p_{bound} vs ϵ for $\frac{N}{V} = 0.25$.

then p_{bound} has some intermediate value. This arises because the terms in our equation reflect a competition between energy and entropy. It is energetically beneficial to bind to the ligand (if $\epsilon < 0$) but taking a

ligand out of solution reduces the solution's entropy. At energy scales around k_BT it is this competition that determines the system's behavior. It turns out that the energy scales of a number of biological processes live right at this border. For these reasons, it is convenient to use k_BT as a ruler for energies. To summarize,

- At energy scales much smaller than k_BT, thermal fluctuations and entropy dominate.
- At energy scales much larger than k_BT, energy minimization dominates.
- At energy scales around k_BT, the competition between energy and entropy determines system behavior.

3.2 DNA Force Extension

The creation of optical tweezers has allowed experimentalists to examine the effect of forces on biological systems. One class of experiments that has arisen from these tools are force extension experiments. A polymer, often DNA, is held in a fixed position at one end. An optical tweezer is then used to pull on the other end with a constant force and the end to end distance of the polymer is measured. This experiment provides a good example of how statistical mechanics can be used to generate a quantitative model.

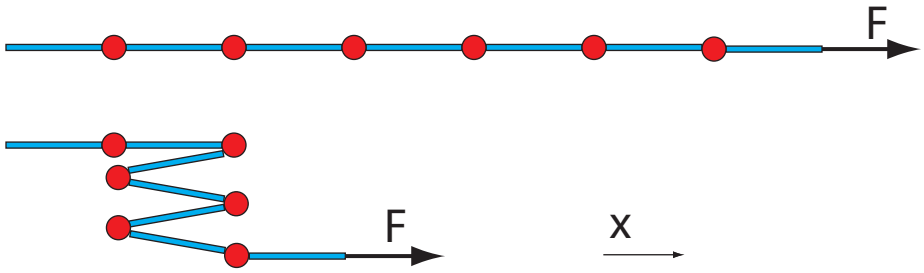


Figure 5: DNA force extension experiment.

For this experiment let's model DNA with a one dimensional random walk. There are N segments of length b and they are allowed to point either right or left. Let N_R be the number of segments that point right and N_L be the number of segments that point left. Naturally we have the constraint

$$N = N_R + N_L. \tag{15}$$

We want to find the average extension as a function of the applied force. Let x be the distance of the end along the x-axis. Then

$$\begin{aligned} x &= (N_R - N_L)\, b, \\ &= (2N_R - N)\, b. \end{aligned}$$

The energy of a configuration is given by

$$\begin{aligned} E &= -Fx, &(16)\\ &= Fb\,(N - 2N_R)\,. &(17) \end{aligned}$$

Before continuing, we need to know the number of microstates correspond to a given extension. The extension is completely determined by N_R. We have N segments, so the number of ways we can choose N_R of them to point right is just $\binom{N}{N_R}$. The partition function is then

$$Z = \sum_{N_R=0}^{N} \binom{N}{N_R} e^{-\beta F b (N-2N_R)}, \tag{18}$$

$$= e^{-\beta FbN} \sum_{N_R=0}^{N} \binom{N}{N_R} \left(e^{2Fb}\right)^{N_R}, \tag{19}$$

$$= e^{-\beta FbN} \left(1 + e^{2\beta Fb}\right)^{N}, \tag{20}$$

$$= \left(e^{\beta Fb} + e^{-\beta Fb}\right)^{N}. \tag{21}$$

The average extension is given by

$$< x >= \sum_{N_R=0}^{N} b\left(2N_R - N\right) \frac{\binom{N}{N_R} e^{-\beta Fb(N-2N_R)}}{Z}. \tag{22}$$

If we stare at our expression for the partition function long enough, we see that if we differentiate Z with respect to βF and divide by Z we get $< x >$. In other words,

$$< x > = \frac{1}{Z} \frac{\partial}{\partial(\beta F)} Z, \tag{23}$$

$$= \frac{\left(be^{\beta Fb} - be^{-\beta Fb}\right) N \left(e^{\beta Fb} + e^{-\beta Fb}\right)^{N-1}}{\left(e^{\beta Fb} + e^{-\beta Fb}\right)^{N}}, \tag{24}$$

$$= \frac{Nb\left(e^{\beta Fb} - e^{-\beta Fb}\right)}{\left(e^{\beta Fb} + e^{-\beta Fb}\right)}, \tag{25}$$

$$= Nb \tanh\left(\beta Fb\right). \tag{26}$$

Again, the results agree with our intuition. With low applied force, entropy wins and the DNA stays crumpled. With high applied force, energy wins and the DNA is elongated.

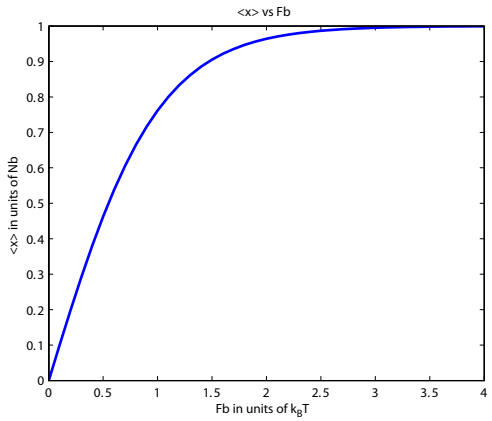


Figure 6: Force-extension curve of DNA.

4 Recommended Reading

- Dill, K. and Bromberg, S., *Molecular Driving Forces.*