Lab exercise 4: Diffusion - Biophysical Chemistry

April 9th 2018

1   Recap

Looking at molecular systems in terms of statistical physics and biophysical chemistry, you will agree that systems tends towards lower energies, sometimes in unexpected ways. In the first exercise we saw this through the enthalpic difference between microstates, and Boltzmann statistics.

<div style="text-align:center; border:1px solid;">

A low energy leads
to a high population.

</div>

Typically we are more interested in general properties than highly specific cases. For instance, **there are many ways to be in the "open state", so we create an "open" macrostate of our own definition, which contains the set of all microstates where the protein behave "open-like".** We then saw, that the volume of its microstates, *the entropy of the macrostate*, directly influenced the (Helmholtz) *free energy* of the state:

$$F = E - TS \quad \text{(Helmholtz free energy)} \tag{1}$$

Viewed as acting upon macrostates, the Boltzmann expression should then use the Free energy F:

$$\langle N_i \rangle = N \cdot \frac{e^{-F_i/kT}}{Z} \tag{2}$$

We were also able to construct an algorithm which reproduced the Boltzmann prediction. At first this might have seemed a pointless restatement, but we could then employ that algorithm to make simulations. Those simulations in turn seemed to indicate that there is a dynamical property of these systems which the Boltzmann equation did not give any information about.

<div style="text-align:center; border:1px solid;">

If there are barriers, these determine
only the dynamics (kinetics) of a system.

</div>

Subsequently, we found a new way to characterize a system of states; how frequently we exchange between specific states. A transition matrix describing all these probabilities thus provides a more complete picture, from which both populations and rates can be extracted.

In todays exercise, we instead ask what dynamics we can expect from a system where there are almost no barriers, and what that can tell us about both our methods (and physics).

2 Introduction

2.1 "directed" processes

Imagine a protein which pumps ions from a low-energy state, across a barrier, to a higher energy state (for instance creating a higher concentration on one side of a membrane). This results in a population which appears to not agree with Boltzmann statistics.

It may seem to defy the Boltzmann equation, but the protein is in fact consuming energy from somewhere else. If we include this "somewhere else", then we can see that Boltzmann is again correct. This is because the Boltzamnn equation applies only to a so-called *closed system*, where nothing is acting upon it from outside. Ions may in this case be pumped into high-energy states by at the same time moving ATP into a lower-energy (hydrolysed) state. The sum of these changes will in fact always show a decrease. So the closed system (with ATP included), *does* obey Boltzmann statistics. The protein in this case simply acted as a catalyst for the following reaction:

$$\text{ATP} + \text{low-energy ion} \rightleftharpoons \text{ADP} + \text{P}_i + \text{high-energy ion} \tag{3}$$
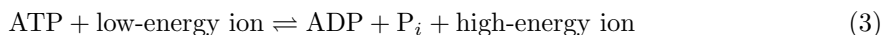
2.2 "undirected" processes

Say that we use the ion-pumping protein from the preceding section. At some point there is no more ATP, and the pumping stops. From what we know about Boltzmann statistics, it seems that if we leave this system for long enough, the barrier won't matter; the ions will re-distribute so that most of them are at the low-energy side of the barrier. How long that takes, depends on the barrier, but eventually it *will* happen. The Boltzmann equation says it has to be so, because once the protein is no longer pumping ions, the ions alone are a closed system.

This thought-experiment shows us something very fundamental. It turns out that processes which are "driven" in one direction are opposed by a "driving force" towards "disorder". Any work we do to defy Boltzmann statistics will be undone eventually. It would be useful to know how strong or fast this "undoing force" is. Diffusion is a direct characterization of this, and this is why we will study it today.
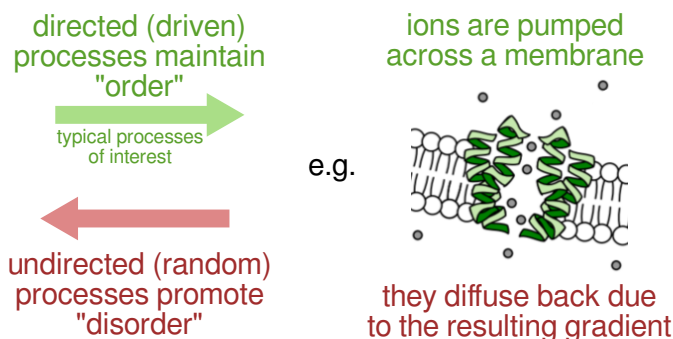
Task

Consider the example of pumping ions above as a closed system. Let's say for the sake of simplicity that there are 100 ions in the system. We chose an ion, an follow it around over time. More precisely, we check every millisecond if the ion has changed moved to the **outside** of the membrane since our last observation. This will count as it having been "pumped".

1. If there is more ATP in the system than the protein needs to pump 100 ions, will the ATP ever be completely depleted?

2. If we observe an ion being pumped (according to the definition we just made), can we be sure that an ATP was consumed?

3. If wait infinitely long, how will the ions be distributed between inside and outside?

Hints

1. You can solve this type of problem by using logic based on what we've learned, by constructing a simple model. For instance; let's say that each ion is in either of four states; "inside" (I), "outside" (O), "transition" (T) or "pump" (P). The energy of state I is slightly lower than state O, but state T is a lot higher in energy than both I and O. P has the same energy as O. You can only get between the inside and outside by going via T or P. You can only get to P from I, and I in turn leads directly to O. The reason we make these rules is because *they make sense in terms of what we described for the ion-pumping system.* Draw this on paper, putting each state at a height corresponding to its energy, and drawing arrows according to the above description. What will happen in such a system over time? What if all ions start on the outside?

3 Free diffusion

So a system will, given enough time, equilibrate to the Boltzmann distribution. It will do so by random fluctuations which eventually reproduce the Boltzmann prediction, since those random fluctuations are modulated by energy-based probabilities. In short; systems appear to want to be maximally random or "disordered", but have to obey energy. Note how similar this is to our simulation algorithm; **chose randomly, and accept based on energy.**

In this exercise we will try to understand and quantify this "disorder". Since we feel like this disorder is independent of the energies which modulate it, we will study it by creating a simulation where all energies are the same. A practical situation which is very similar to this is for instance a single water molecule in solution; it does not appear to settle (get trapped) in any low-energy state, but simply bounces around continually. The aim now is to derive some characteristic of this type of process, in our case by making simulations and drawing conclusions from the results. For this we make a few modifications to our simulation model:

- We will **not** need to calculate the Boltzmann factor to decide if we go or stay, because all energies are the same:

$$P(A \rightarrow B) = e^{-\Delta E/kT} = e^{-(E_i - E_i)/kT} = e^{-0} = 1 \tag{4}$$

We will therefore simply take randomly chosen steps between states.

- We will not define states explicitly, but rather say that each unique position is a state. If we are at position 3, we are in state 3, for instance.

- We will not define an explicit topology, but rather say that we can only access adjacent states (those next to us). If we are in state 3, then we will only permit going to state 2 or 4, for instance.

This will create a so-called random walk. It exemplifies a process known as *free diffusion*. Diffusion models how random fluctuations are expected to accumulate, which perfectly matches our

needs.

Now let's actually construct a free-diffusion model of a particle moving in space, based on the above. Our first model will have states at each integer position along the y-axis, and will start at y=0. At each simulation-step we then move either 1 step up or 1 step down, selected randomly. This is then a "1D random walk", and it will behave as illustrated in the figure below.
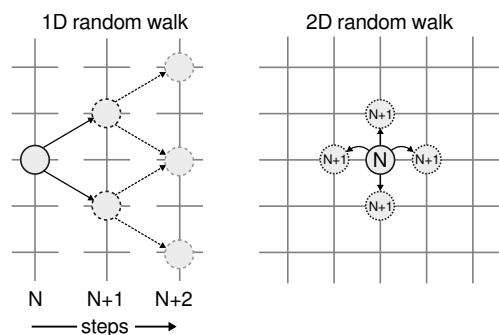

Figure 1: In a random walk, we pick one of the adjacent states with equal probability, and move there. We never reject a step, because this is *free* diffusion, where there is no Boltzmann factor to reject a move, nor are there any barriers.

Tasks

1. Open the file lab4_1.py , and edit line 17 to make a random suggestion at each step of the simulation, suggesting either a step up or a step down. Show line 17 in python code.

2. Run the file and see where you end up, the plot will show the position as a function of time (the number of steps taken). Include a plot.

3. In our previous exercises, our simulations performed two actions: randomized a suggested state, and accepted/rejected based on energy. Is this a random walk? Argue your choice (use at most 2 sentences).

Hints

1. `np.sign(`x`)` returns -1 if $x < 0$ and $+1$ otherwise.

4   Characterizing diffusion

We will now try to understand something about how random walks "behave", and what we will arrive at will help us understand a wider range of diffusive processes. Ideally, one would like to be able to arrive at a statement like

Proposition 1:
"If I start at 0 and take N steps, then I will end up at position Y".

and then also determine how Y depends on N. If we think about what a "random walk" means, we quickly realize that this will probably be impossible. Imagine we close our eyes and walk around a unknown city for some time to see where we end up. We do this twice, and both times end up in the exact same place. Clearly it is more likely that we were somehow led there, either by the city's design, our own behavior, or something entirely different. Critically, this is hardly a "random" walk.

So, **what can be expected of a random walk, if not the position at the end of a walk?** The second file lab4_2.py makes a number of random walks, according to your settings, and then plots a number of characteristics.

Task

1. Use lab4_2.py to make a predictive statement about random walks in general, by changing the parameters and making predictions about what should happen. Ideally you will be able to produce a mathematical statement supported by your collected data, but this is not strictly necessary. What we are asking you for is a simply an observation or rule you believe to be true. It does not have to be correct, but it should be supported by evidence.

Hints

1. You can set the number of steps to be a random number in some range, on lines 49 & 50. This might be useful to gauge how results vary with increased length of walks.

5   Mathematical formulations

Let's try quantify some observations, and try to come up with a measure of diffusion which we can see in real-life experiments. Let's make a new proposition based on what we have seen so far;

Proposition 2:
After N steps of a random walk, the average distance d reached is proportional to the square root of N.

If we express this by a formula through an expectation value[1], and massage it a bit, it looks like this;

$$\langle d^2 \rangle = K_N \cdot N, \quad \text{where } K_N \text{ is some constant we don't know anything about at this point} \quad (5)$$

Task

1. Does this form look familiar too you? What is it? Check the hint if you do not recognize it.

2. The file Lab4_2.py is in fact equipped to use the above answer to estimate K_N. Set "switch" on line 118 to "True" and run the script again. Show a plot and comment on the result.

Hints

1. Try setting $\langle d^2 \rangle = y$ and N = x. What does equation (5) look like?

6   Mathematical interpretation

Now, the water-molecules in a glass of water do not really perform steps of displacement, but rather float around as a function of *time*. So a more realistic formula for diffusion should be

$$\langle x^2 \rangle = K_t \cdot t \tag{6}$$

But even so, we need to think about what K_t *means.*

---

[1] the average or mean expected after repeating the measurement infinitely many times, denoted here by $\langle \rangle$

Task

1. What unit does K_t have?

2. If K_t is increased, how will the system behavior change?

3. Determine K_t for some combination of settings. Then change the step-length L on line 33 in Lab4_2.py, and determine K_t again. Relate K_t to the step length L, and use it to improve equation (5). What is the new equation?

7   Confinement - finite size simulations

If we have moderately realistic systems with many particles (like a protein in water), we can't allow particles to escape away to infinity, like our random-walker does at present. We need some way to keep them within a reasonable distance.

Tasks

1. Modify the script to somehow (there are multiple "correct" ways) keep a simulation within the range of [-10, 10]. Show the code that does this.

2. Estimate K_t again, under the same conditions as you did before. Look at the plots and judge if the above task caused any issues with this estimation method.

In general, this is a common issue in performing numerical simulations to replicate real-life phenomena; understanding how the model and its output are affected by our necessity to compromise in the computational details. We *can* always fit a parameter to data, the question is if that parameter is meaningful or correctly estimated. As you can see, it is perfectly possible to extract a value for diffusion from a system were it is incorrectly modeled.

8   2D-diffusion

It turns out that equation (6) is fairly general. If one starts from very general theoretical foundations, then there is a partial differential equation which describes heat propagation and diffusion (heat also diffuses), with a fundamental constant D;

$$\frac{\partial f}{\partial t} = D \cdot \nabla^2 f \quad \text{(heat equation)} \tag{7}$$

One can solve this equation with random walks in mind, and what then emerges is an equation on the form

$$\langle d^2 \rangle = \langle x^2 \rangle = 2 \cdot D \cdot t \quad \text{(1-D diffusion equation)} \tag{8}$$

Using equation (3) for a 2D-random walk, we will find that the equation seems slightly different, but importantly with the same value for D;

$$\langle d^2 \rangle = \langle x^2 + y^2 \rangle = 4 \cdot D \cdot t \quad \text{(2-D diffusion equation)} \tag{9}$$

The value D is normally called the diffusion constant, and it is different from what we assumed by a factor 2 (that is, $K_t = 2 \cdot D$) in each dimension. This factor should be included if we want to have the diffusion constant to have the same physical meaning as the theoretical foundation of

equation (3), so in practice we always include it. However, even using K_t as a measure of diffusion is completely sufficient to quantify and compare diffusive phenomena.

Tasks

1. Give an example of 2D diffusion in biophysical systems

2. Give an example of 3D diffusion in biophysical systems

3. Let's say we build a simulation model. We run it, and find that it reproduces a wide range of experimental results very well. However, when we calculate the diffusion coefficient of water in our simulation model, it's much higher than what we measure experimentally. Can you use this model for research?

9   Extras

9.1   More about random walks

If you want to see just how common random walks are in building models in science, feel free to check out https://en.wikipedia.org/wiki/Random_walk#Applications

9.2   Mathematical justification

Lets say that the random walk takes a more general form, where a step of length δ cm is taken every τ seconds. Then clearly after a single step past some time t,

$$x(t + \tau) = x(t) \pm \delta \tag{10}$$

so squaring this;

$$x^2(t + \tau) = (x(t) \pm \delta)^2 \tag{11}$$
$$= x^2(t) \pm 2\delta x(t) + \delta^2 \tag{12}$$

Now when we take the expectation of this, the middle term cancels[2] due to the $\pm$. So,

$$\langle x^2(t) \rangle = \langle x^2(t - \tau) \rangle + \langle \pm 2\delta x(t) \rangle + \langle \delta^2 \rangle \tag{13}$$
$$= \langle x^2(t - \tau) \rangle + \delta^2 \tag{14}$$

No we know that $x(0) = 0$, so clearly after the first step, $\langle x^2(\tau) \rangle = \delta^2$. In fact, looking at equation (14), we can see that each step adds another δ^2 to the expected value of the squared distance. This leads us to formulate the hypothesis that if the total time t is the sum of N steps ($t = N \cdot \tau$), then

$$\langle x^2(N) \rangle = N \cdot \delta^2 \quad \Leftrightarrow \tag{15}$$
$$\langle x^2(t) \rangle = \frac{t}{\tau} \cdot \delta^2 = \frac{\delta^2}{\tau} \cdot t \tag{16}$$

Equation (15) is exactly the form of the earlier equation (5), and shows that the arbitrary constant D is in fact proportional to the step-length squared, as we observed in our model. In deed equation (16) is also on the form as equation (6), showing that if each step takes less time, then the displacement is greater, all perfectly reasonable.

There is a very concise and elegant way of deriving the general form of the diffusion equations, which starts from an entirely different point of view. If you are interested, feel free to check it out; http://nebula.physics.uakron.edu/dept/faculty/jutta/modeling/diff_eqn.pdf

---

[2] For a single random walk this is not true, but we are talking about the expectation value, which considers an infinite set of random walks. This is the same thing as considering an infinite set of equations like (12), of which half will have $+2\delta x(t)$, and the other half will have $-2\delta x(t)$, so the expectation value becomes 0 since these cancel.