Lab exercise 7: Free energy landscapes - Biophysical Chemistry

April 19th 2018

1 Introduction

The general approach so far has been to build models which replicate aspects of the real world, and then using those models to get a deeper or better understanding of precisely why things behave or happen the way they do. In today's exercise we will show how our models can be used in an additional capacity, as general tools employed as a part in solving a specific problem.

Let's say we are trying to develop a new drug and have a number of potential candidates. We need to determine how well each of them binds to the receptor we want to affect. This is very similar to what we did in a previous exercise. In that exercise, we simply "asked a program" to figure out how well each drug bound to the protein. The aim today is to understand how a program is able to do this. What we formally want then, is to evaluate how favourable a state P is over another state Q. What we then seek is the *free-energy difference* between states P and Q, often called ΔG. For clarity, we can also (in this case) say Δ_{PQ}

In the present case, we let the state P represent the drug in solution outside the receptor, and state Q be the drug bound to the receptor, as shown in figure 1. We might also be interested in how easy it is for the drug to cross a membrane, in which case the states P and Q will represent the drug outside and within a membrane.

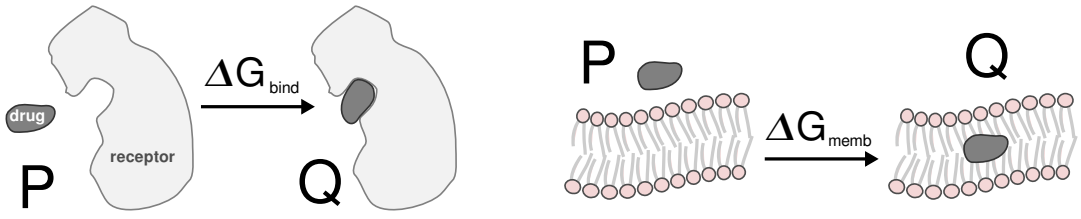

Figure 1: State P specifies the substrate in solution, while Q is the substrate in a bound state in the left figure and embedded within the membrane in the right figure.

1.1 Tasks

1. What does a negative value of ΔG_{bind} mean on the left in figure 1?
2. What sign does ΔG_{memb} have on the right in figure 1, if the substance finds the membrane to be a barrier?
3. Generally speaking will the substrate ever cross the membrane if ΔG_{memb} is positive?

2 The Entire Path

In this exercise we will examine what the challenges are to determine ΔG, and examine one strategy to meet that challenge using simulations.

Let's examine a different problem which will be useful to think about later. You are standing in a valley, on one side of a mountain. You now want to know if the valley on the other side of the mountain is higher or lower than the one you are currently in. **How do you find this out?**

While there are many clever ways of doing this, the important lesson for todays methods, is this: **If we know the profile of the entire path (as in figure 2A) between the two valleys, we can measure or calculate the total difference.**

One way to establish this path is to simply walk across the mountain, and every 1km we stop and measure how steep the incline is. After arriving in the other valley, we draw lines of equal length end-to-end on a piece of paper, each one with the angle we measured that part of our walk. This re-creates an approximate profile of the mountain as shown in fig 2B, and we can then measure on the paper what the total difference is. Mathematically, there is a much simpler way[1], but the principle is the same. Let's examine figure 2 more closely. Because we chose segments which

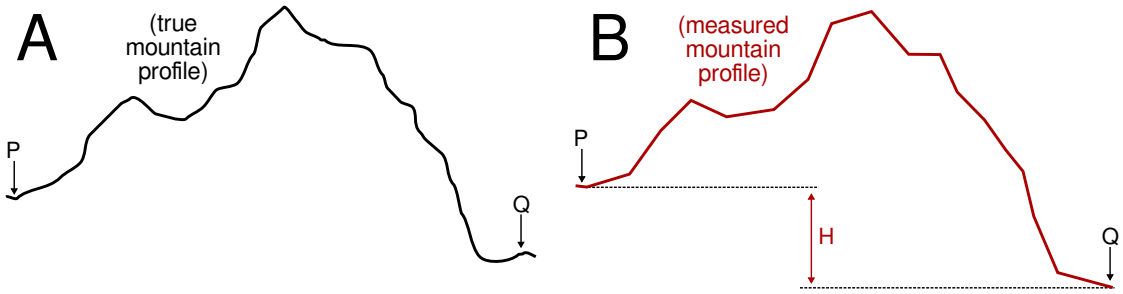

Figure 2: We can estimate the difference in height between two points P and Q if we have an estimation of the entire path between points P and Q.

were quite long and perhaps our measurements of the incline had some errors, the estimated height difference was slightly (or very) wrong, but in general we can always reduce the length of the intervals and improve the measurements of the incline. In mathematical principle, this is a general procedure which will always work. How well it will work, is a different matter altogether.

2.1 Tasks

1. If you were to actually measure the angle d (in degrees) of a mountain hillside every x meters, decreasing the interval x between measurements will increase the accuracy of the height measure H (see figure 2B). True or false?

2. In the same example, you estimate the total path across the mountain is ~10km. What value for x would you choose if I asked you to actually do this with a real mountain? We're asking you to assess and decide, not to be correct.

[1] sum the sines of the inclines and muliply by 1km.

3 Missing a Part of the Path

Let's say that we walked across the mountain, collecting *very* detailed data *very* often. However, there was a region called "Silent Hill", and sadly the data for this region has been lost. This has been sketched in figure 3.

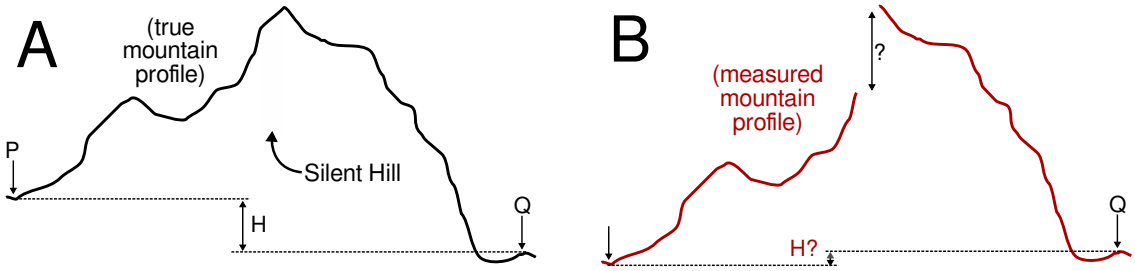

Figure 3: The people you sent to gather data from the "Silent Hill" region didn't return. Do we care?

Hopefully you will agree that unless we have a *complete* path from P to Q where we know the incline at effectively every point, there is no way to estimate the total difference in height with any level of accuracy. *Even if we know all other points perfectly.* The same is true for calculating values of ΔG for molecular phenomena.

We often speak of a "free-energy landscape" because it is easy to think of going between states P and Q (like in figure 1) as traversing some landscape. But in our simulations, we find jumps between these valleys, and rely on high sampling to count the populations and invert those to find energies and energy-differences. But some barriers are so high that it would be extremely time-consuming to await this sampling. But we don't expect to observe high-energy states either, so estimation of the landscape would require a more general way to target sampling to a certain region.

3.1 Task

1. There might be more than one way (path) to get from state P to state Q. For the height measurement H (as in figure 2B), it **does not** matter which such path you chose to measure. True or false?

2. The height measurement H, **does** become more or less expensive to compute depending on the path chosen. True or false?

3. How accurate the height measurement H is, **does not** depend on the path chosen. True or false?

4 The Incline of the Free-energy Landscape (in Simulations)

There is a specific picture which we have now draw for you many times; the energies of a few states and the resulting distribution according to Boltzmann-statistics, as shown in fig 4A.

We have used this simple illustration to explain how the population of a number of (discrete) states relates to their energies. Now we can generalize this to a *population density*, typically

referred to by the Greek letter ρ [rho]. In figure 4B we have illustrated this by making an energy that is a function of the position x, and showing a qualitative idea of what $\rho(x)$ can be expected to look like.

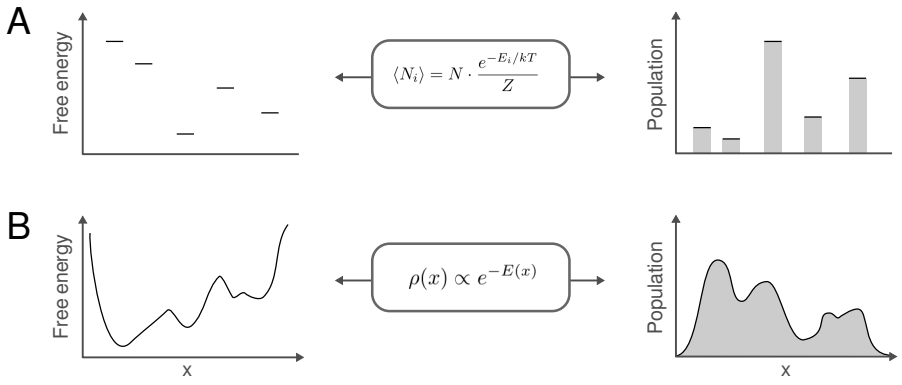

Figure 4: Boltzmann statistics for a continuous variable x. NOTE that energies in A are NOT related to enrgies in B.

4.1 Task

1. We have made an intentional error in figure 4. What is it?

5 Sketching the energies and populations

Hopefully the previous exercises have given you some intuition or gut feeling for how populations will look given the energies of a system of states. This will be useful as we now try to "trick" our simulations into showing us the incline of the free-energy landscape wherever we want to know it.

Take the system in figure 5A, where we have a single positive charge P which is free to move along x in a tunnel. A little outside the tunnel we have placed other charges, which cannot move. The point P cannot go outside the tunnel.

There is an additional example 5B, where there are no external charges, but springs which force the particle to obey Hooke's law:

$$F_s(x) = -K \cdot x \tag{1}$$

where K is a spring force-constant.

5.1 Task

1. For figure 5A, sketch the free-energy landscape and population you would expect to see if you could simulate the particle P according to Boltzmann-statistics (yes, you should draw the graphs). This does not have to be analytically correct, you are only required to make it *qualitatively* correct. If you do not know how to draw on the computer in a simple way, I suggest this site. you can even "import" the image files from the mondo resources into that editor and simply draw directly on them.

2. As above, for figure 5B.

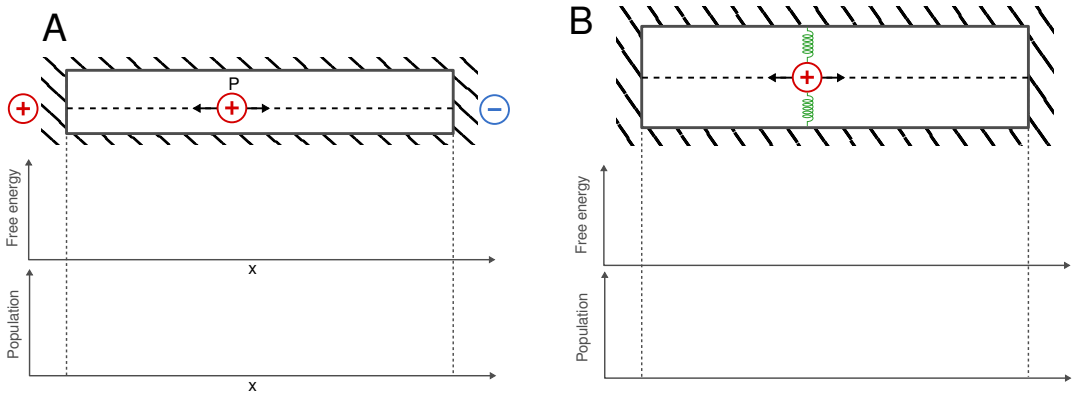

Figure 5: Sketch $E_c(x)$ ('c' for charge) for the charged particle P, then sketch $\rho(x)$. Then do the same for $E_s(x)$ ('s' for spring) in B, and do note that there are no external charges in system B.

3. Use equation (1) to create an expression for the energy of the particle P in fig 5B. Then plug this expression into the formula in figure 4B, what does the resulting density-expression $\rho(x)$ look like?

Hints

1. For task 3, integrate equation (1) from 0 to x. This will give you the total force required to move the point P from 0 to x.

6 Combined

Now lets consider a system where both the external charges and springs act on the charge P. This is shown in figure 6.

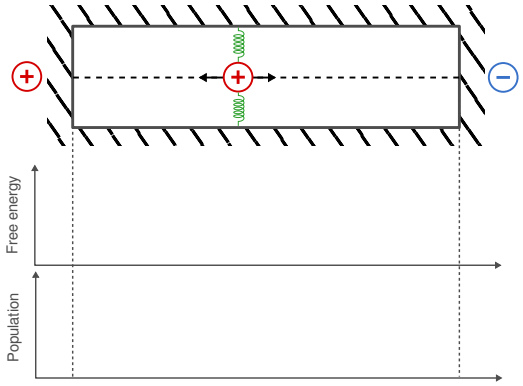

Figure 6: Sketch $E(x)$ for the charge P, then sketch $\rho(x)$.

6.1 Task

1. As before, draw the energy and populations you expect to find, this time for figure 6.
2. Looking at your sketch, can you think of a way to calculate the contribution from the external charges $E_c(x)$ if you know the spring force constant(s)?

7 A more realistic system

Let us for a moment think about the mountain we had to cross in the first part of this exercise. The conclusion was that we had to sample *all* the points along the path across the mountain. If we simulate a system, this is equally true; *if we want the ΔG of two states, we have to adequately sample all points along the path connecting them.* Look at the system in figure 7. Again we have a charge P which is free to move along x, and which is affected by a number of external charges. By now, you know what we want you to do, but this time it is not so clear what the ΔG will be...

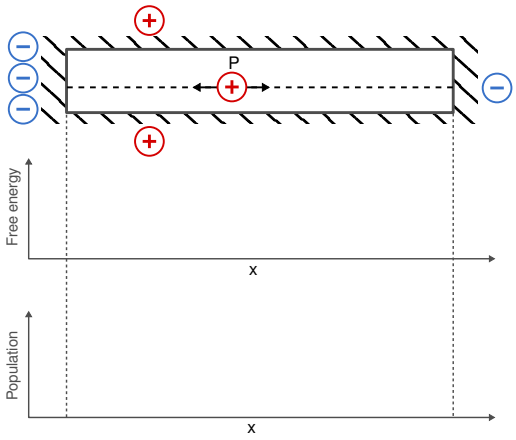

Figure 7: Sketch $E(x)$ for the charge P, then sketch $\rho(x)$. Which side of the positive charges if "better" for P to be? The left side is more negative, but closer to the positive external charges.

7.1 Task

1. As before, draw the energy and populations you expect to find, this time for figure 7.
2. If we tried to simulate this system, we might find that the barrier is so high that we never go across, even with extremely long simulations. How would you force the charge P to be in high-energy states and still find the incline of the energy landscape you sketched in figure 7?

8 Umbrella sampling

The method of using artificial but *known* forces to keep a particle in a certain state, to find the slope of the free-energy landscape, is very commonly used. Even if a state is really favourable and definitely would be seen during a simulation, it is a really good way to *target* sampling to interesting regions, and reduce the amount of time we need to simulate systems to examine the aspects we want to observe and quantify.

Umbrella sampling is *not* a method which performs a single simulation, but rather an ensemble of many simulations that are used in a collective analysis. In each simulation the system is restrained to a particular position along the path between two states, so that together all simulations cover the entire path. This is very similar to measuring how steep a mountain is at many different points.

We have processed the output from a set of such simulations, which explores the path of propane crossing a membrane of POPC lipids. By running the script `plot_1.py`, you can visualize the sampling of the landscape, and the landscape itself.

8.1 Tasks

1. Should the profile be symmetrical?
2. Inspect the simulation coverage. Is there anything strange about it?
3. Run `plot_2.py`. What's different now and what is the effect on the energy landscape?

9 Conclusion

We started out by realizing that we have to have sufficient knowledge of the path connecting states, in order to establish or estimate the difference in energy between them. We knew that this would be difficult to overcome, because those states tend to be connected by high-energy barrier which are hard to sample in simulation. We then realized that we could restrain our simulation on a specific way, and simply subtract or account for that biasing force, to estimate the underlying landscape. Making this a systematic method, like umbrella sampling, allowed us to sample even high-energy paths sufficiently to find the the difference in energy between states. We still don't know everything about the system though, as we have only seen one out of possibly very many ways to go from P to Q, but the energy difference is at least possible to find in many cases, since we only need one path, regardless of which.