Lab exercise 3: Kinetics - Biophysical Chemistry

March 31st, 2016

1 Introduction

By performing Monte Carlo simulations on an extremely simplified model using states with fixed enthalpies, we have been able to explore multiple fundamental phenomena present in real biophysical systems. We were also able to see that if we group multiple states together into *macrostates*, then those macrostates are associated not just with an enthalpy, but with an entropy as well.

In this assignment we will return to this model once more, this time to study the effects of applying constraints on the system. This will introduce the concept of having a barrier between two states, in addition to just a the difference in enthalpy and entropy between them. We will use a similar python-application as in Lab 2, with a similar command-line interface and input/output-file procedure, but with a slightly modified simulation algorithm.

Download the supplementary files for this assignment and place them in your local computer. To make sure the simulation engine is working properly try running

```
./mcrun2.py -h
```

If everything is working this should print out the command-line instructions, listing all the application options. **Create an enthalpy-file called enth.dat, containing four (4) lines, each having a number representing the enthalpy of a single state. A good range is between -200 and 0. Run a simulation and verify that the results are reasonable.**

2 Creating a Network Topology for "a system of states"

So far we have been simulating a system of states, where all microstate-to-microstate transitions are possible: We can go to any state in one move, *regardless of which state we are currently in.*

A real biophysical system *is* however subject to "constraints". Bonds and other interactions within the system will for instance limits its mobility, making certain transitions between states less likely or practically impossible. As an example, the open and closed state of a protein may be separated by physical distance; jumping instantaneously from one to the other is so unlikely that we might as well cal it a *forbidden transition.* In order to go from open to closed, the protein has to go through one or more so-called *transition states.*

To be able to study such effects and how they impact the dynamical properties of a system using our simple model, we need to be able to define which transitions are allowed, and which are forbidden. In terms of a connected network of states, this means we have to specify the

connections, or define the system topology.

In the previous exercise we used flags like "`-t 300`" to specify which temperature to run the simulation at. This time, we will use the additional flag "`-net topology.dat`". If this flag is set, the application will look for the file specified. If the file is found, it will use it. If the file was not found, then a file will instead be written out, which you can modify and use as later input. Try running

```
./mcrun2.py enth.dat -net topology.dat
```

Where "`enth.dat`" is your enthalpy-file. If you didn't have a file named "`topology.dat`" in the working directory, you should have one now. It should contain four lines and four columns, all being set to 1. **Check that this has happened**. You can use this file to define a system topology, by making certain transitions forbidden. Here is how this file should be interpreted:

Each row corresponds to the state the system is going from,
and each column represents the state the system is going to.
For every simulation step, a suggested transition form state A to state B
considers the value at row A of column B.
If this value is "1" the transition becomes subject to the Boltzmann factor as before.
If this value is "0", we stay in state A, regardless of energies of A and B.
Hence, a "0" signifies a forbidden transition.
Figure 1 shows two examples of this, for a simple system like yours.

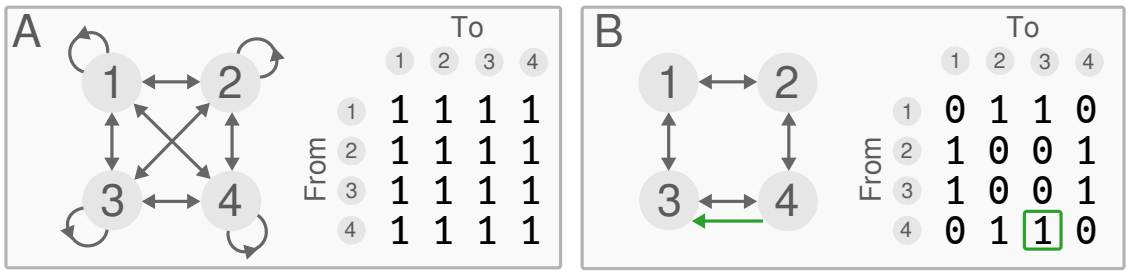


Figure 1: Microstates 1,2,3 and 4. Topology represented by arrows for allowed transitions. In case A, all transitions are permitted, just like in the previous exercise. In B, some transitions are forbidden. This is reflected in their topology, which represented by the same 16 digits contained in the "topology.dat" file, where 1 means "permitted", and 0 means "forbidden". An example of this is shown by a green arrow, which has the corresponding digit circled, also in green.

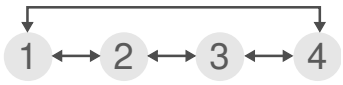
Tasks

1. According to the Boltzmann equation, knowing the energy of our four states completely determines the population of each state. True or False?

2. The energy of state A and state B is NOT changed by making the transition between them forbidden. True or False?

3. Should the population of states change when a single transition is changed from "permitted" to "forbidden"?

4. Write down the digit-topology (the 4-by-4 matrix) of this system of states:



Hints

1. For task 4, you can count the number of arrow**heads** in the figure, the total number of ones should be equal to this number.

3 Simulation

Let's use this new tool to run a simple simulation. **Create a topology.dat-file containing the topology from task 2.4. Also, Create an enth.dat-file containing 4 enthalpies in the range between -50 and 0.**

The aim now is to see what effect forbidden transitions will have on the simulation and the results we derive from that simulation. Run a simulation using the file you just created:

```
./mcrun2.py enth.dat -net topology.dat -n 10000
```

Plot the distribution of states using the "plot.py" file:

```
./plot.py
```

Tasks

1. By running simulations with different number of steps, try to estimate the number of steps required to consistently get roughly the predicted proportions of state populations (a rough number based on inspecting the plots by eye is fine.)

2. Now alter the topology.dat-file, to make the transition between state 1 and 4 forbidden in **both directions**. Repeat the previous task for this topology.

4 A trapped simulation

In realistic systems, there are always a few states which are much more stable than most others. These are called called *stable states*, since we will likely stay there for quite some time. We might even say that the system is "trapped" in such states. We might also call it a meta-stable state, depending on how long we stay in it, but this is not a strict definition.

But now we are referring to a span of time and how long we tend to stay in a state... the Boltzmann equation makes no statements about time-dependent behavior of a system of states,

so what is going on? Precisely what determines for how long we are "trapped" and what kind of sampling we need to have to be confident about our simulation?

Let's try to exaggerate this time-dependent effect to understand it better. We do this by creating a some extremely stable states which are only connected by a very bad state, at least an order of magnitude higher compared to k_BT. **Use the topology from task 2.4, and change the enth.dat-file to look like this:**

```
0
-2100
0
-2300
```

Tasks

1. Run for 10.000 steps and plot the distribution. Do you observe "trapping"? Show the plot.
2. Use the flag `-i` to set the initial state of the simulation (by default the simulation starts form state 0). Does the plot look different depending on where the simulation starts?

By now it has hopefully become clear that the required sampling depends heavily on the topology of the system being sampled, and the transition states and barriers that become inevitable because of it. The time-dependent behavior then seems to be related to the barrier connecting states.

> Forbidden transitions can disable direct connections between low-energy states.
> This forces the system to deal with energy barriers.
> By defining a topology we are able to see the effect of energy-barriers.

5 Kinetics

We began by defining a system of states, and asking the question "How will this system be affected, if we make certain transitions forbidden?". The key to answering this question lies in the fact that transitions are time-dependent phenomena, and the Boltzmann equation makes no statements about time. An answer seems to be that "We change the time-dependent dynamics by forcing the system to change in certain, specific ways, but the population is unaffected." This is what is known as kinetic control. A system which is *not* free to move however it wants to will still obey Boltzmann statistics, because the Boltzmann equation does not care about time.

Task

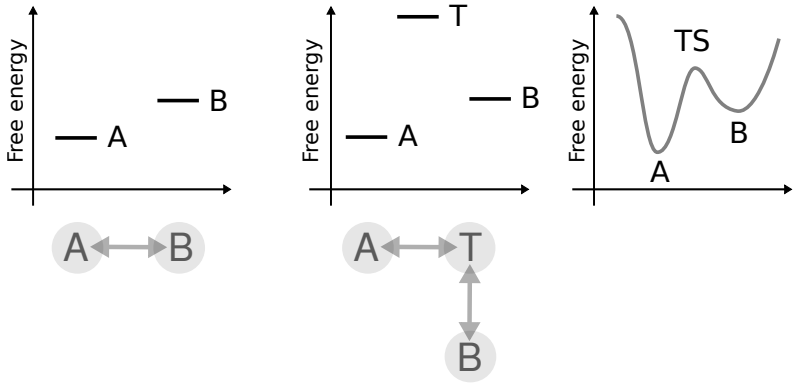
1. How long do we have to simulate any system to get results in *exact* agreement with the Boltzmann equation?
2. Think about a general reaction, where two substances collectively go from being in the state "separate" to being in the state "combined". How does a catalyst change the population of these states?

3. So what does a catalyst do?

Protein folding it starts off with a random coil, which explores the free-energy landscape and starts to approach the native (folded state) state. This proceeds quite quickly, much quicker than simply trying all configurations until the best one is found. This indicates that there is an order to how it proceeds. This in turn indicates that there are some meta-stable states which have allowed transition connecting them, but that virtually all other states and transitions are unfavorable. Such *folding intermediates* and the barriers connecting them are said to exert *kinetic* control over the folding process. The rate, and even a probable sequence of events during folding, can be maintained by this mechanism. This is exactly what we have concluded above; the energies determine the populations, but how those populations evolve over time depends largely on the height of the barriers connecting them.

6 Transition state theory

There is a way to unify the concepts we've been discussing so far, and it is known as transition state theory. Let's forget for a second about "forbidden transitions". We again have a model of 4 states, each associated with an enthalpy. We perform a simulation just like before; suggest a random new state and decide to accept/reject based on some consideration of energy. But instead of asking for the relative energy between source and target states, we define a "transition state" which has some energy. This energy is effectively a barrier height. The higher the barrier energy, the more unlikely it is that we will go to the top of it (i.e. cross it).



In conclusion, kinetics are dynamical properties of systems which are vital to understand actions, rather than just states. Barriers are the fundamental reasons we observe kinetic control and controlled reaction phenomena. Because we want to understand dynamics, not just states, we would really like to see the transition state.

Task

1. A transition state is rarely observed. Why?
2. Give a (very) brief example of a biological phenomena where a barrier is lowered.

Below follows extra material for those who are curious about how to derive the distribution directly from counting transitions, rather than counting how often we are in states.

7 Estimation of Distribution directly from transitions

There are also ways to improve the approximation of the distribution using only poorly sampled simulation data. A relatively simple approach is to look at how the equilibrium distribution relates to the transition probability. For a simple two state system we know form *detailed balance* that the following holds

$$\pi_0 P_{0,1} = \pi_1 P_{1,0} \tag{1}$$

Where π_i is the population of state i and $P_{i,j}$ is the transition rate form state i to j. Noting that $P_{0,1} = (1 - P_{0,0})$ we get

$$\pi_0(1 - P_{0,0}) = \pi_1 P_{1,0} \quad \Rightarrow \quad \pi_0 = \pi_0 P_{0,0} + \pi_1 P_{1,0} \tag{2}$$

It can be shown that this relation expanded to N states becomes

$$\pi_i = \sum_j^N \pi_j P_{j,i} \tag{3}$$

Using this equation you can try to improve the approximations of the state populations, π_i. The file "plot_approx.py" tries to do the same as the previous plot-file but instead of plainly counting the states in the trajectory-file it counts the transitions between the states and uses a variation of equation (3) to approximate the state distribution.