20.320, notes for 11/13

Tuesday, November 13, 2012
9:42 AM

Last time:

Today:

1. Refining structures in pyrosetta
2. ΔΔG for small molecules
3. Drugs!

We have done lots of approximations in order to get reasonable calculations

- Only a small change in the protein's backbone
- We only consider rotamers
    - We work with a discrete set of X angles

Now we work on refinements to those first approximations

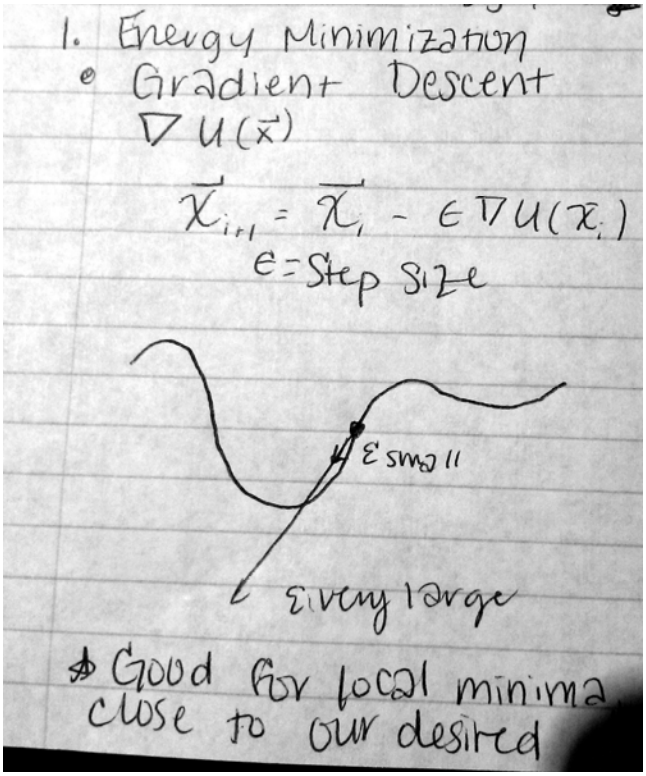
1. Energy minimization
2. Molecular dynamics

We spent a lot of time figuring out how to calculate overall potential energy as a function of either rotamer angles or x-y-z coordinates. At the end of our algorithm, given our approximations, we will most likely not get to the lowest possible free energy state. We aren't considering all rotamer angles (only discrete steps), so we're likely to not be able to get to the perfect bottom of the energy well.

In mathematics we could use the first derivative of a function to see whether we were at a minimum. For a grand multi-dimensional space we could similarly calculate the gradient. This works great for smooth functions, but not for our messy, empirical relationships. Instead, for our real case we use something called **gradient descent**.

Gradient Descent

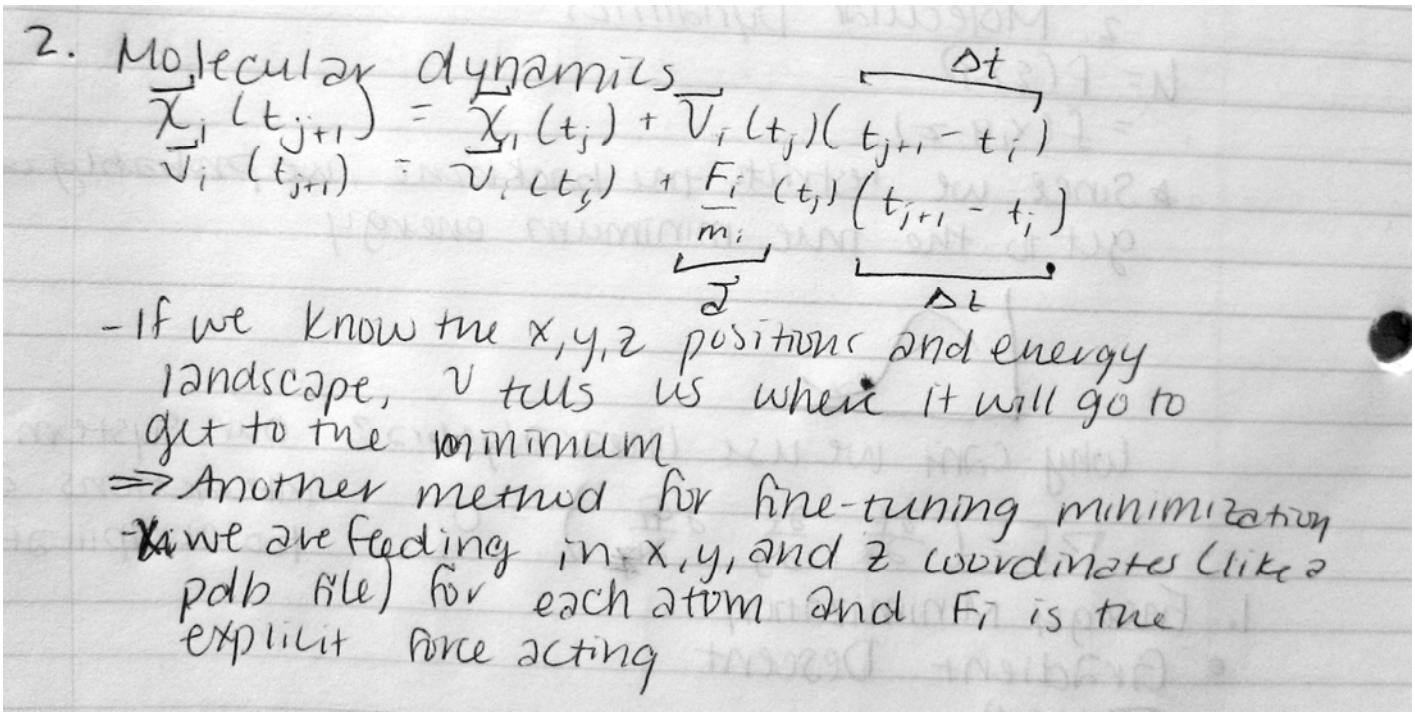
At any point on the energy landscape, we can compute the gradient in all directions and determine which way it goes down. When we have that, we simply move in that direction. Note that we've used the metropolis algorithm with our blunt approximations to do the first sampling of the space. It can search an enormous space, with simplifying constraints determined by us, and will hopefully reach the global minimum. It will get us close to the ideal, we hope, at which point gradient descent can fine-tune our final values.



Molecular Dynamics

For this method, we start assuming that forces (from molecular interactions) act on every atom and affect their motion. For each atom we can calculate its current location and figure out where it will be given its velocity and the acceleration caused by the forces around it, over a small time step. Atoms will move until they reach the state of lowest energy within a well.

Note that both gradient descent and molecular dynamics are unable to jump over an energy barrier to reach a lower global minimum elsewhere. They are always stuck at whichever local minimum they start out at. That's why we run the Metropolis Algorithm first.



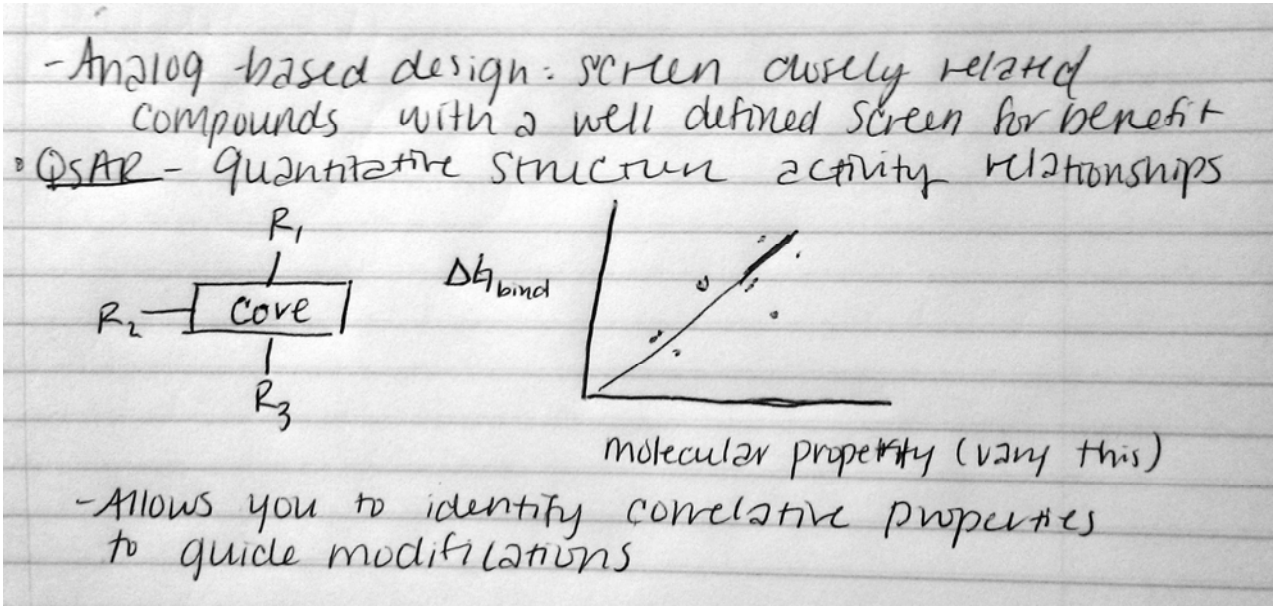
ΔΔG for small molecules

Let's shift our focus to the sort of substances that most drugs are: small molecules. How do we apply these concepts we've learned? For that we need to look at how drugs are designed. There are 2 ways:

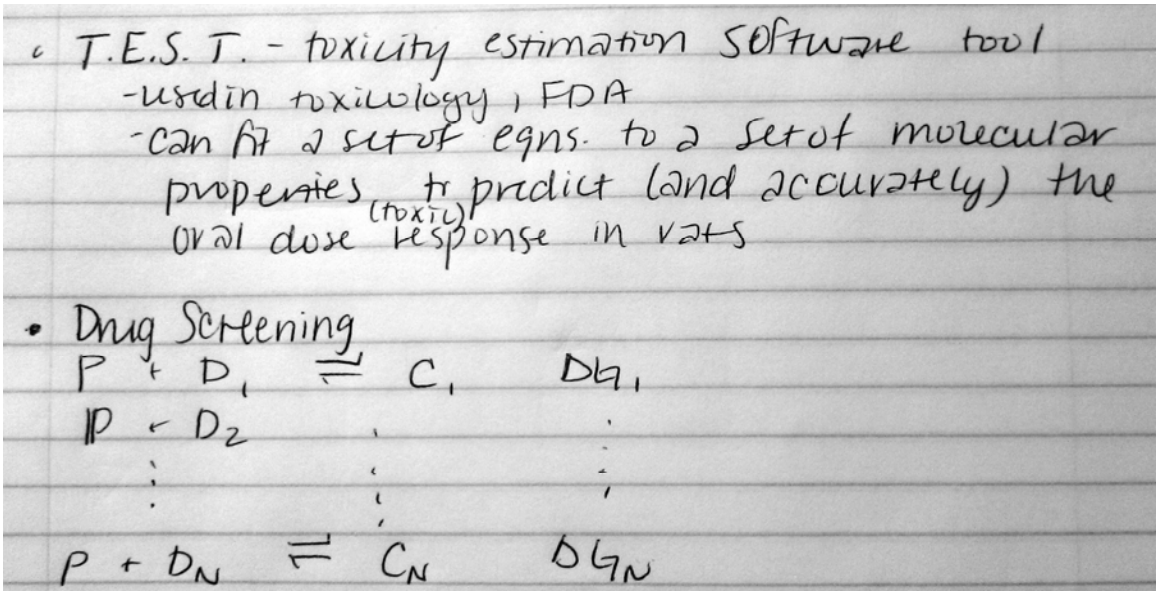
Analog-based design

It's been around for thousands of years. It was the only way before computer simulations. The idea is to take a starting material known to have a therapeutic effect (like salicylic acid). You try a whole lot of chemical modifications (like acetylation, among many), and you assay for effectiveness until you find the best (acetylsalicylic acid, Aspirin).

The most rigorous and modern way of doing this is through a method called QSAR (Quantitative Structure Activity Relationships). The idea is also to change stuff randomly onto a starting molecule, but then you try to see whether there's a pattern in the data. Is there any characteristic of our molecules that seems to correlate with greater effectiveness? If so, we try to maximize more of that.



There is also a popular equivalent in the toxicology world, where they use TEST (Toxicity Estimation Software Tool) to determine the toxicity of stuff by essentially the same method.



Structure-based design

All this looks very similar to protein design. They also calculate free energy diagrams and what not. However, there are some important differences. The search space for drugs is much, much greater than the already astronomical search space for proteins.

Proteins	**Drugs**
Limited number of variables (20aa)	An estimated 10^{60} compounds that would make realistic drug candidates
Mostly fixed backbone	Nope
Known active site to mess with	Active site(s) unknown

So imagine we want to see where a small molecule binds to a given protein. The way to do it is to search around the protein, dock the drug somewhere, and create a score for that binding spot. The scoring, of course, comes from the same tools that we'd used before. Calculate which spot shows the lowest potential energy of interaction.

To do our search, we need six independent degrees of freedom. Three are for location (x, y, z) and three for orientation (the equivalents of roll, pitch, and yaw, to use the aerospace analogy). This is enough to describe all potential interaction sites. As for how to sample them, we again bring in our method for sampling enormous sample spaces: the Metropolis Algorithm.

Does this work? Well, it certainly reduces the search space. It is good at determining the really bad binders, even if it has more difficulty telling apart the good ones. Just knowing what to ignore, though, is already a great starting point for the pharmaceutical industry. You find leads and then try to improve on them. The leads can be analyzed by X-ray crystallography or NMR to see exactly how the drug in fact binds to the protein.

Pharma priorities

The pharma industry lore has created the labels of good targets and good drugs.

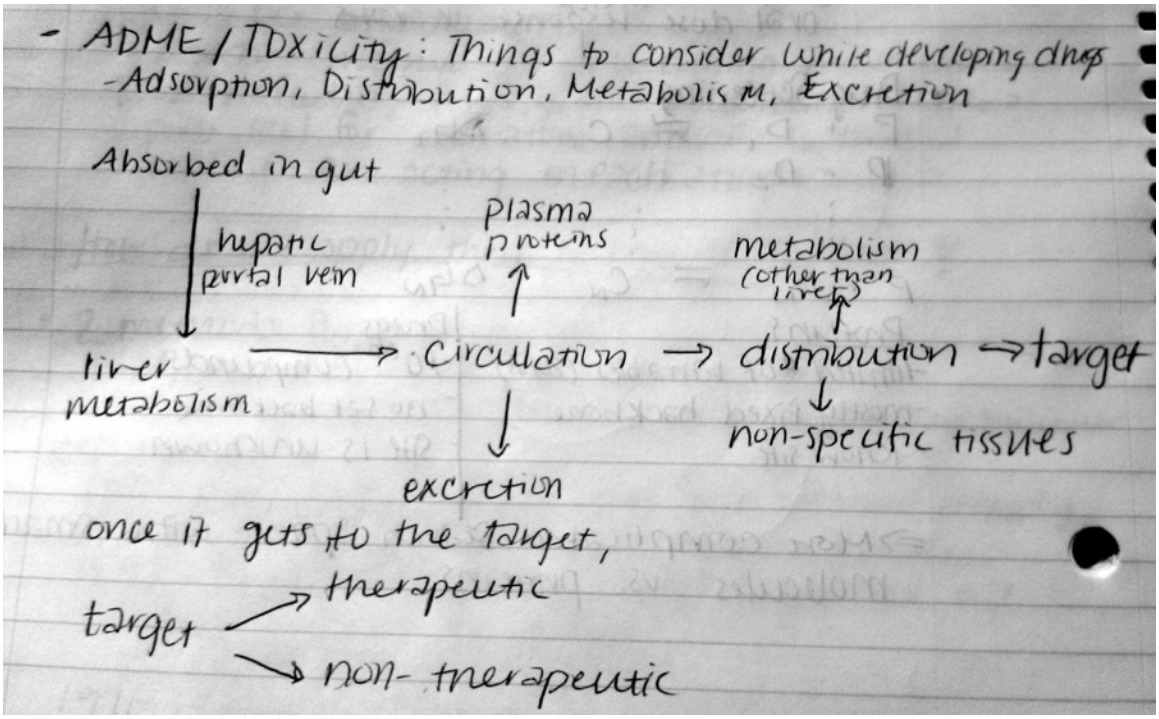
Good targets are those that are likely to be a monetary hit. They have made lists of what they call

druggable and undruggable targets. The druggable ones are those for which we already have several examples of working drugs (kinases, GPCRs, Ion channels) that we can use as starting material. The undruggables are those that nobody has tried, or that many have tried and failed at.

More important for our discussion, though, is what makes a good **drug**. How do you sift through 10^{60} candidates? Well, there are considerations other than the molecular interactions. This is where we look at ADME/Toxicity, the short list of considerations that will make or break a potential drug:

- Adsorption
- Distribution
- Metabolism
- Excretion

Most drugs are small molecules taken orally. Their life cycle inside the body looks something like this:



So when the industry looks at a molecule, it needs to consider all of the above. How do they do it? Well, they rely on lots of calculations and a few famous rules of thumb. One of the most famous was put together by a guy called Lipinsky, based on his experience of what has and hasn't worked in the history of the pharma industry. It is a short list of the things that a successful drug should **not** have.

Lipinsky's Rule of Five

A good drug candidate should **NOT** have:

- More than 5 H-bond donors
- Molecular weight > 500 Da
- Sum of N, O atoms > 10
- High lipophilicity, defined as below:

$$log\left[\frac{D_{octanol}}{D_{water}}\right] > 5$$

- Also avoid an area > 140 square angstroms

- Avoid > 1 formal charge

20.320 Analysis of Biomolecular and Cellular Systems
Fall 2012