

20.320, first class

Thursday, September 06, 2012
9:30 AM

Today in 320

Goals and rationale for the class, followed by some intro material.

Professors: Forest White and Ernest Fraenkel.

TAs: Allison, Thomas, Daniel

Intro

The goal of this course is to teach you to **model biological systems** at different scales. We want predictive and quantitative models. You've all seen the cartoon models in peer-reviewed papers. This class is about how to turn that cartoon into equations capable of predicting the behavior of the system. The first half of the class is about the networks of biomolecules. In the second half of the class, we look at the way the structures of proteins and other molecules arise from their sequence and how this relates to sequence. Ultimately, the goal is to have something predictive enough to intelligently guide experimental design.

And so, we'll cover a variety of topics:

1. Modeling techniques
2. Measurement techniques
3. Manipulation strategies
4. Data mining

Admin details

There is no book. It is only your mind that bends.

There are some optional readings and suggested books. Old problems and exams are all posted on the website. The best way to prepare for our exams is, of course, to work through old problems.

Exams

The dates of the two exams are listed on syllabus. Oct 4th, Nov 15th. Each worth 20% of the grade. Final is cumulative. There's also 2 design projects, as opposed to the one project that there used to be. One is due halfway through, on kinetic modeling. The second is about protein structure, mutations thereof, etc. They're both essentially week-long take-home exams, which means that there is no collaboration allowed. You can, however, come to the profs and TAs for pointers and assessment of your selected topics.

This year, there are no problem sets. *Vox populi, vox dei*. That time has now been put into the design projects, and we hope that this won't crash and burn. We will hand out problem sets; they won't be graded, but it's in your best interest to ignore that fact.

Grade distribution

We use the full range of grades, from A to F. Historically (3yrs), 20% have dropped or earned a D in the class. There is no quota; there is no allotted number of failing grades that we will offer, and you shouldn't think there is. Work hard, get help, come to class and recitation. And just try. We use the MIT

definitions for grades.

By drop date, 60% of the material will be on record. We will still, however, not be able to predict your final grade. Students with a potential A can still bomb the class. Students with potential C's can still turn it very far upward. The TAs are *not allowed to tell you how you're doing in the class*. They will, if anything, refer you to the profs for guidance.

Design projects

We'll ask you to pick a system and design a quantitative model around them. Say, a synthetic beta cell. Imagine you had a simplified system designed to have some essential components that you'd normally find in a beta cell. Design it all *in silico*. This is a chance to use your creativity. Design all the way from receptor to transcription and expression. When you code a working model, based on realistic mechanism, you can then analyze the robustness and sensitivity of your model cell. Hopefully, determine the important nodes of the network. This is a way to do *in silico* synthetic biology.

The second project is to take a bromodomain (or a similar protein domain). The goal of this project is to look at protein domains and see how they interact with other molecules. These two projects will allow you to go in the lab and predict the changes that are likely to have a desired effect.

Recitation and lectures

There will be two recitations. A Doodle poll will help determine the best times for our two recitations. Go to at least one of them per week, please, though they are optional.

Software

The first half is about MATLAB. There will be a tutorial. This is the standard tool of computational biology and systems of ordinary differential equations. The second half will use python and Py-Rosetta.

More detail about class content

Section 1: Modeling complex dynamic biological systems

How they work, how they respond to perturbation. The primary example of this throughout the class will be the MAPK (Mitogen-Activated Protein Kinase) cascade. Kinases are enzymes that phosphorylate other proteins or themselves, turning ATP to ADP and putting the phosphate onto the protein.

The kinase cascade is a cascade of kinases. It starts at the top, with MAP kinase kinase kinase. We're not making this up. This is the main signaling cascade, from yeast to human, that is used to respond to exogenous signals.

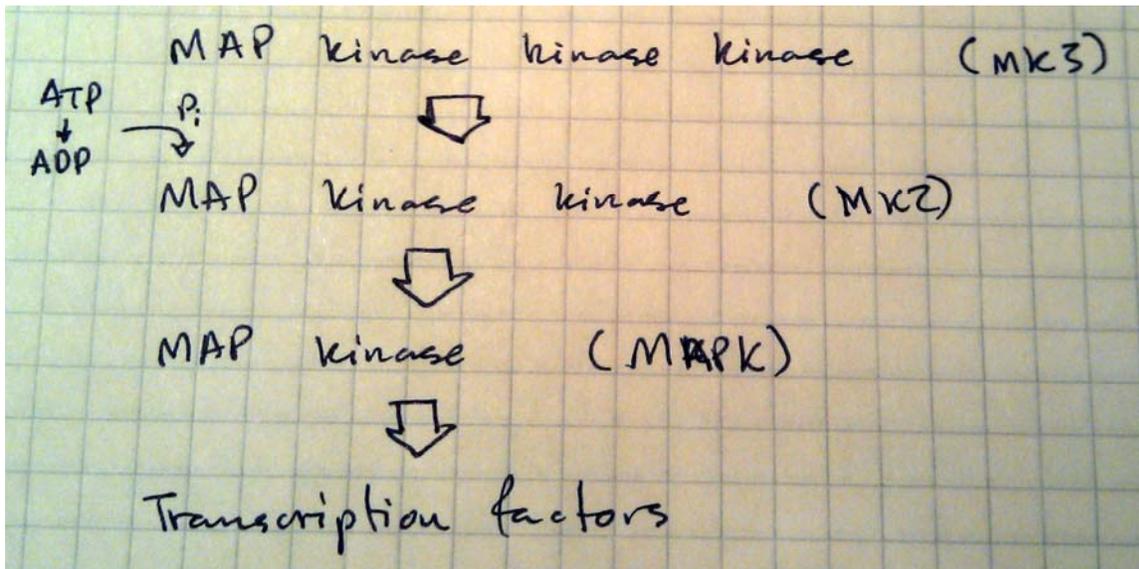


Fig 1: The MAPK cascade

In vivo, things are a tad more complicated. There's a scaffold protein in yeast called Sterile 5, which holds together all the elements of the cascade in the right orientation and proximity. This increases the overall rate of phosphorylation.

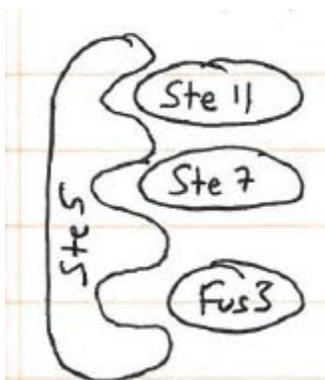


Fig. 2: The MAPK cascade in yeast. (Ste = Sterile)

Cascade elements.

| | |
|--------|------|
| Ste 11 | MK3 |
| Ste 7 | MK2 |
| Fus 3 | MAPK |

What happens when you knock down Ste5? You still get phosphorylation, but at a decreased rate. If you can knock down one of them to a particular extent, which extent is enough to create an important effect? At first, nobody knows. The cartoon cannot tell us. *However*, our quantitative model will indeed be able to give us a numerical answer. That's what we're trying to achieve.

The model can show us the sort of things that are difficult to predict otherwise. If you overexpress the scaffold to great excess, will you just wind up with a whole lot of partially unoccupied scaffold molecules? If you were just doing straight-up biology, you wouldn't be able to interpret the weird results that you are sure to get.

Section 2: Predicting and messing with the relation between protein structure and function

So from the first half of the class you know that you want to increase the affinity of the scaffold for the first kinase. You know which perturbation of the network will have the desired downstream effect. Now you learn how to do it. Which mutations will lead to which effect?

First material: Protein-Ligand interactions

Look at the relation between Ste11 and Ste5. In the 1st half of the class we will generalize that to all proteins (P) and ligands (L). What do we need to know about the system to parameterize our model? What do we care about this interaction?

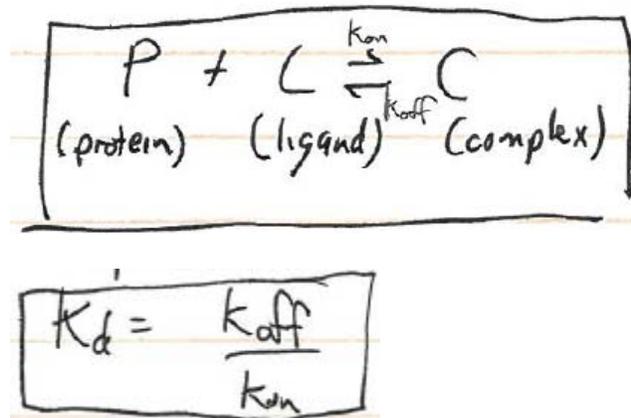


Fig. 3: The basic protein-ligand model.

Furthermore,

$$K_a = \frac{1}{K_D}$$

Now, if we wanted to analyze the energetics of the system, we'd run into this diagram.

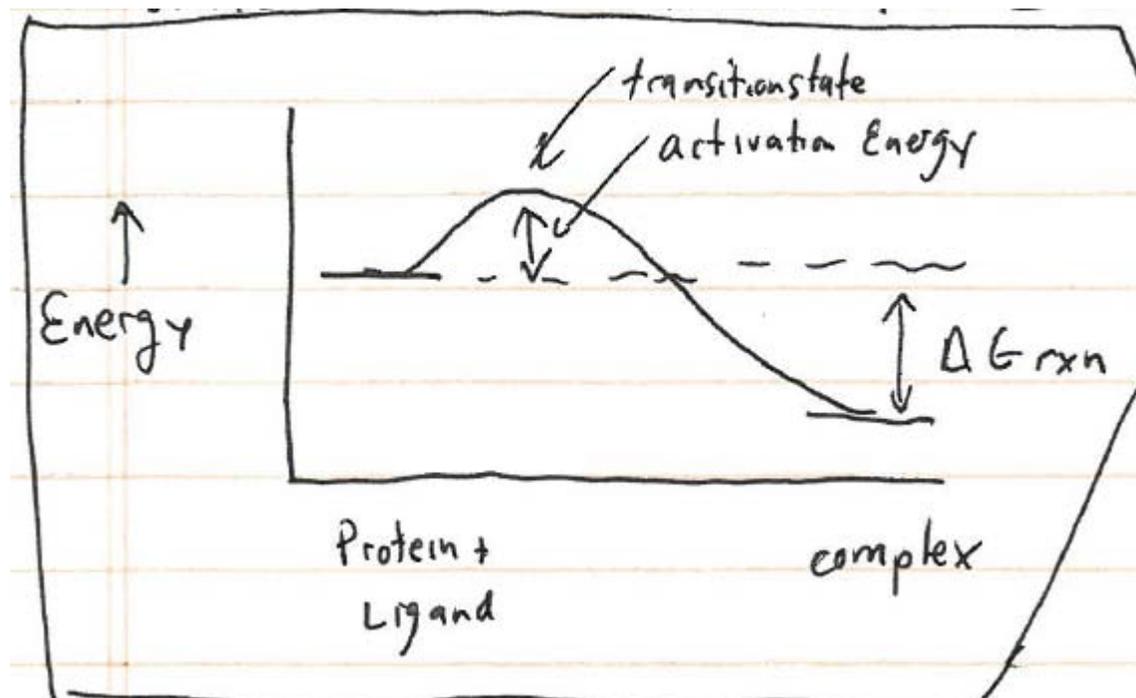


Fig. 4: Energy landscape for protein-ligand interactions.

But what is ΔG in the first place? (It means Gibbs free energy, but we need more than the name).

$$\Delta G_{diss} = RT \ln K_d$$

Where

$$\Delta G_{diss} = \Delta G_{products} - \Delta G_{reactants}$$

And also remember the thermodynamic description of ΔG .

$$\Delta G = \Delta H - T\Delta S$$

Why do we care? Because this equation is a description of the exchanges of energy and entropy of the reaction. It can help you design (or understand) the rates of reaction, of dissociation, and which particular amino acids are responsible for these effects.

And in order to analyze this we need some really good experimental techniques. But that... we'll cover next class.

MIT OpenCourseWare
<http://ocw.mit.edu>

20.320 Analysis of Biomolecular and Cellular Systems
Fall 2012

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.