

20.320, notes for 12/4

Tuesday, December 04, 2012
9:40 AM

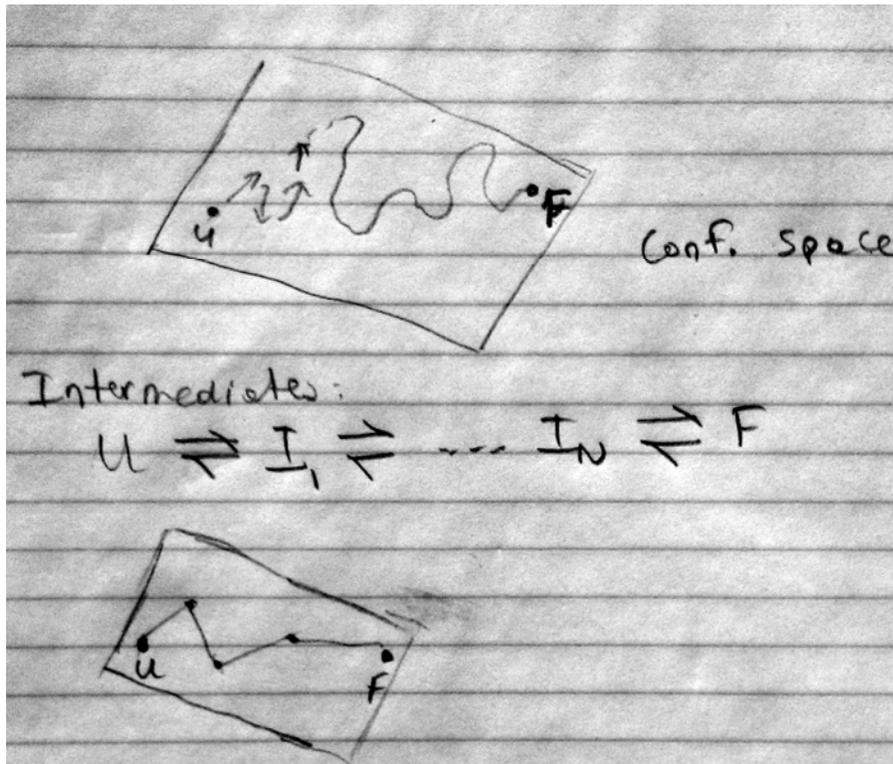
Today

Kinetics of folding
Predicting structure

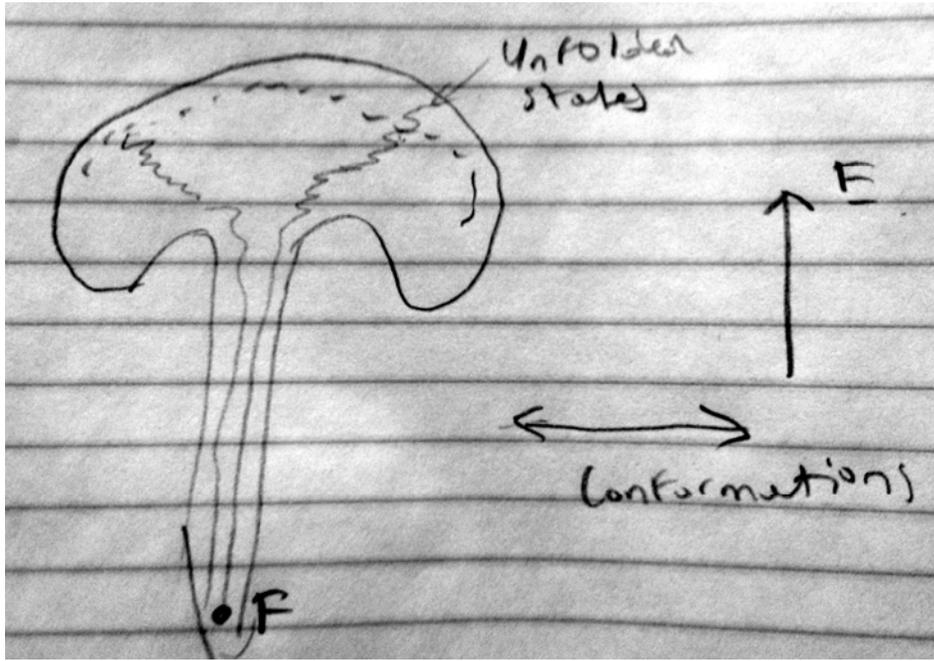
Folding Kinetics

Let's start with a question. How fast does a protein fold? Without making assumptions, let's look at this from first principles. Cy Levinthal, here at MIT, addressed that question once. Imagine a protein with 300aa. If there were three states per amino acid (roughly the three favorable regions in the Ramachandran plot) and the amino acids were sampling through them randomly, there'd be about 3^{300} states to sample, or 10^{143} . At one state per picosecond, that would be 10^{123} years to fold. Since folding really happens in the second to millisecond scale, proteins are clearly not doing that. This came to be called the Levinthal Paradox. What's the method for non-random sampling?

One theory is that the protein goes through intermediate states. There's a whole bunch of them between the unfolded and folded states in the conformational space. There's a folding pathway, in other words, that makes the random walk much, much shorter. The problem with this theory is testing it. The folding intermediates are fleeting (milliseconds, remember?), so they could not really be caught experimentally. The field tried for a long time, but could not produce evidence for them.



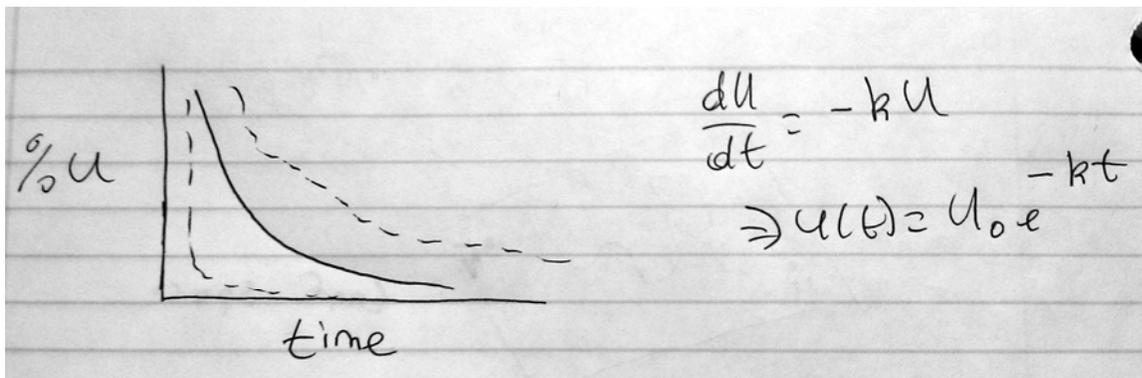
Let's take a more nuanced view at the problem. We know what the folded state looks like, but which is the unfolded state? In reality, we'd expect there to be lots and lots of them. And one thing we've ignored so far is energy. The landscape, when we put energy in the Z axis, looks like a funnel. An energy funnel, with folded state at the bottom and unfolded states up along the rim. This changes things, because every random search is now directed. They are always falling downhill, which makes the pathway much easier even as it continues to be probabilistic.



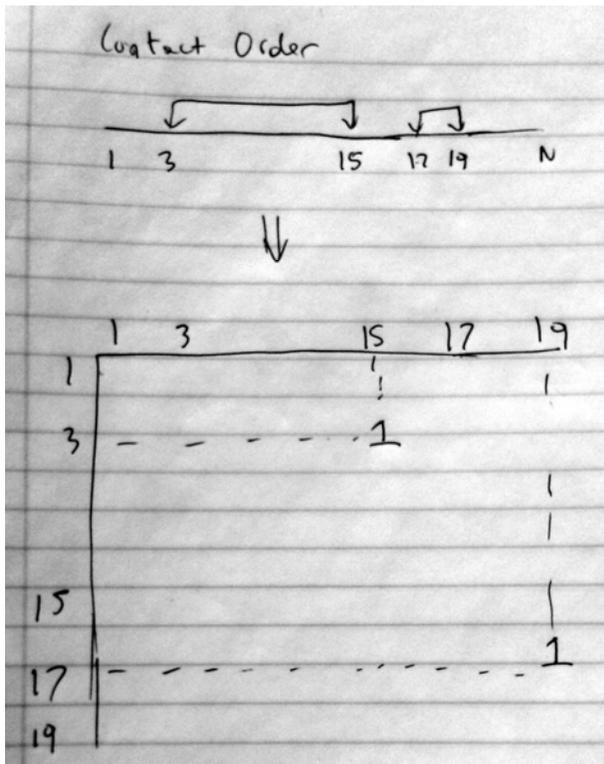
This is a funnel.

Note, btw, that even the folded state is really a bunch of them at once. Thermal noise will always mean some jiggling around. The amount of variation, though, is much smaller than for the unfolded states.

Imagine you take a lot of protein, denature it in urea solution, and then dilute it in water. As the urea dilutes away, it will start folding. If it's monomeric and sufficiently dilute, then it will interact only with itself. The percentage of folded protein will follow a simple exponential decay.



What determines the constant k ? We'll go through a long-winded explanation. People like to think of something called contact order, which is a measure of whether most of the amino acid interactions are local or distant. To determine, it they draw something called a contact matrix, which is a plot of all the physical contacts between the amino acids in the protein.



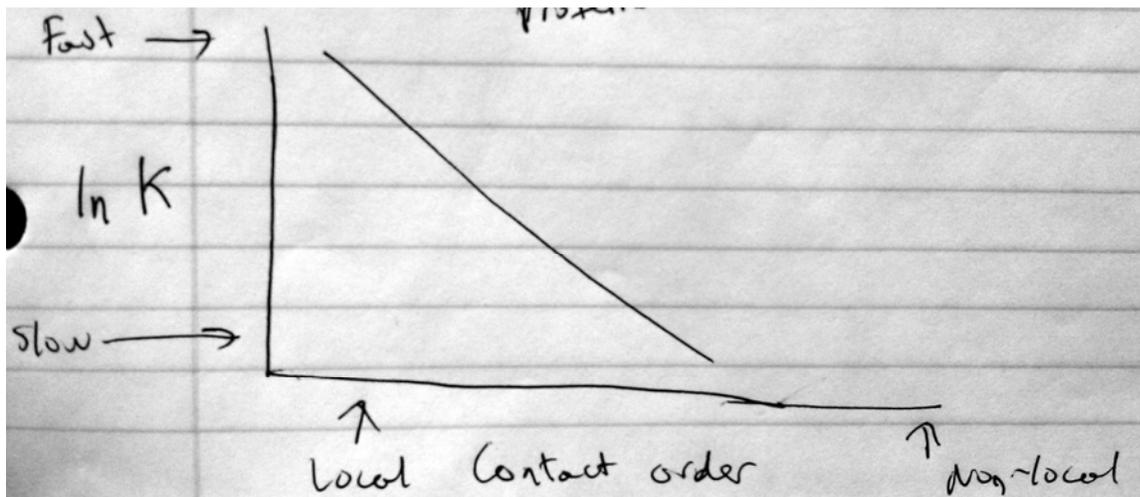
Contacts near the diagonal are local; they happen between close neighbors. We can calculate a single number that represents the degree to which this happens, and we call that the contact order.

$$\Delta S_{ij} = |i - j|$$

$$\text{Contact order} = \frac{1}{L \cdot N} \sum_{\{i, j\}_{\text{contacts}}} \Delta S_{ij}$$

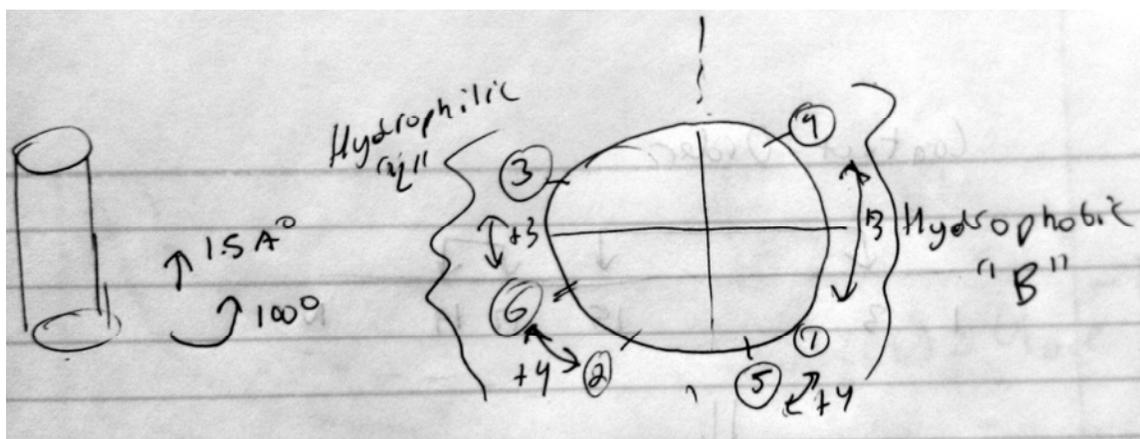
L is length of protein, N is number of contacts.

If we now draw a graph of folding speed versus contact order, what would we expect? More local interactions means a faster-folding protein. It makes sense, as local structures can happen more readily through random interactions. They should take less time.

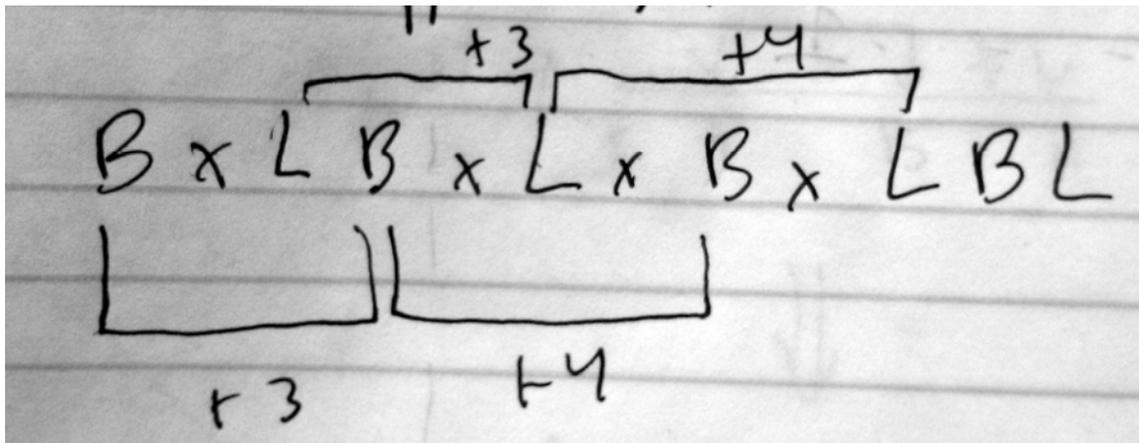


Predicting Secondary structure

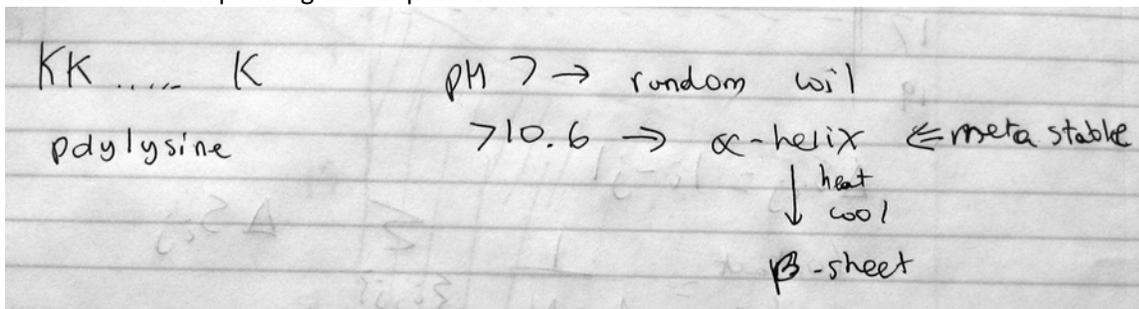
Alpha-helices are the simplest structure to predict. The first one to be predicted was insulin, back in 1967. All you need are a few assumptions about the inside being non-polar and the outside being polar. You can get extra clues if your helix is amphipathic (that is, if one side is hydrophilic and the other hydrophobic). In that case, due to the dimensions of an alpha helix, each amino acid is about 100° around the circle from the previous.



The upshot of this is that amino acids located about 3 or 4 spots from each other tend to be on the same half of the helix. THUS, if you see alternations of hydrophilic and hydrophobic amino acids in a primary structure, separated by 3 and 4 spots, you can guess that you found an amphipathic alpha-helix.



Does this seem awfully limited? Well, yeah. It's only good for finding amphipathic helices. Luckily, there's other knowledge-based algorithms for predicting structure. One of the first (but still best) was one developed by Chou and Fasman. They discovered that poly-lysine has an alpha-helical propensity. At pH 7 it will form a random coil, but at pH above 10.6 it will be metastable, alternating between alpha-helix and beta sheet depending on temperature.



They decided, then, to calculate the propensity of all amino acid polymers (made out of a single aa) to form alpha-helices or beta sheets. They measured the abundance of all amino acids in the known structures of the time, and they classified them as helix formers, helix breakers, or intermediates. They combined this helical propensity with the theory that helices form by nucleation (that is, that a small amount of helix forms spontaneously and precipitates the helixation of the structure around it). They thus created a small algorithm for predicting alpha-helices from primary structure:

1. Find region with 4-6 helix-forming residues (H, h in their nomenclature)
2. Extend in both directions 4aa, and stop if you hit b, B (helix breakers)
3. No proline

This algorithm manages to predict helices with ~60% accuracy. We've only barely been able to improve on that slightly, even after throwing massive computer power at the problem.

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>.