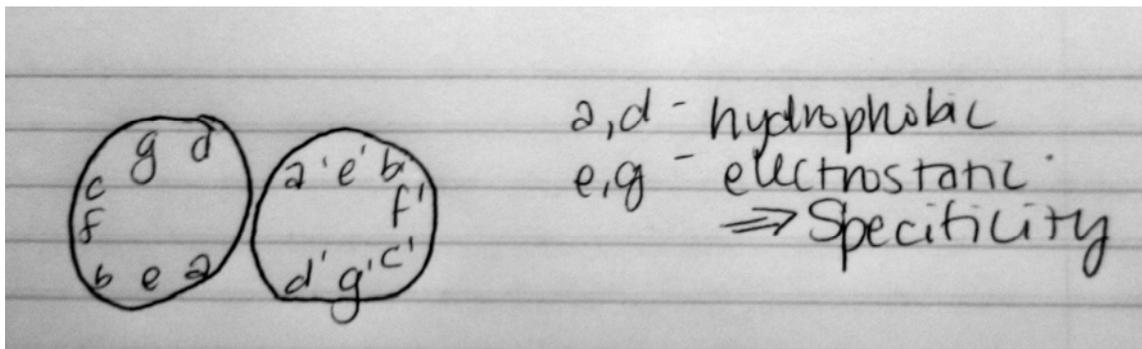
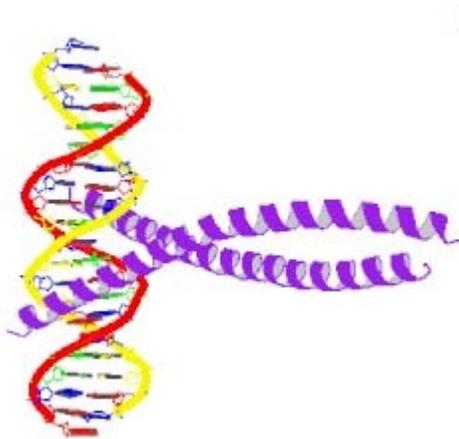


# 20.320, notes for 11/1

Thursday, November 01, 2012  
9:40 AM

## Specificity

Last time we started talking about prot-prot interaction specificity, and we used the example of leucine zippers. If we mark the amino acids around the alpha-helix, we notice the interactions that hold it together.



Cross-section of the alpha-helices, showing the characteristics of the residues around it.

When the alpha-helices of leucine zippers are monomeric, they lack all the stabilization forces from interacting with the other helix. Therefore, they're not even entirely folded up. Their structure is induced by the dimerization itself. It turns out that this is part of a long continuum, that goes all the way from induced structure to rigid docking. Let's look at rigid first. As we've seen before, water will have to be excluded from the interface. Since all those solvent interactions are broken, we need to make up for it with other favorable interactions. This can be achieved through complementary electrostatic charges between protein and ligand.

### Rigid Docking

- ~800 angstroms
- 20 amino acids
- 100 atoms
- Solvent excluded from the interface
- Tight packing around the edges, creating a virtual O-ring.
- Complementary charges

How can we analyze the interactions at this interface? Through a tool called **Alanine scanning**. The idea

is to mutate each amino acid, in turn, to a tiny amino acid that is as non-functional as possible. Alanine is the smallest AA other than glycine (which is too flexible and confounds our data more than it helps), thus lacking side chains that are likely to be significantly polar, nonpolar, sterically hindering, etc.

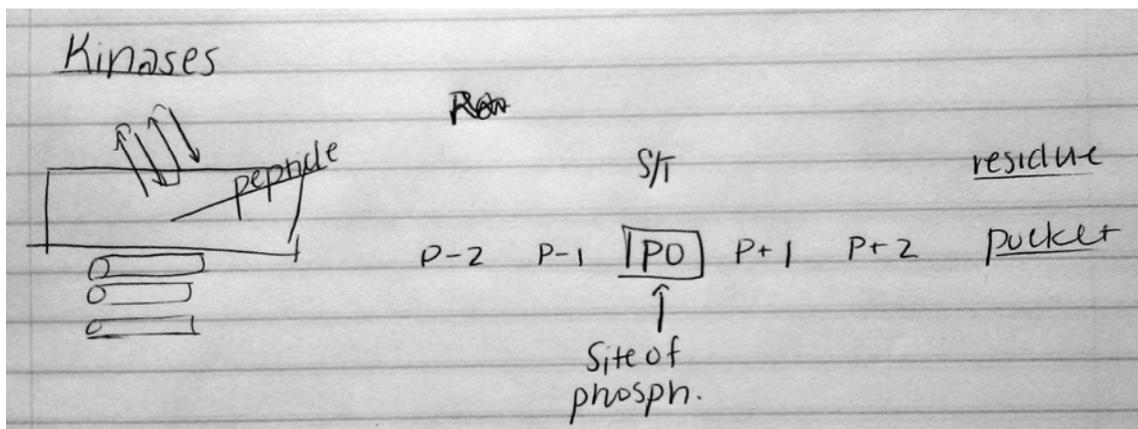
### Alanine scan

1. Systematically mutate single amino acids to alanine
2. Measure Kd's, determining which mutations increase or decrease specificity

Alanine scanning studies have shown that less than 10% of the residues contribute more than 2 kcal/mol to the interaction. They also show that buried charges have a big contribution to specificity.

Let's look at kinases as an example. If we look at commercially available kinases, we notice that they don't just phosphorylate any target. They recognize particular sequences. Examples:

S/T Kinase	Recognition sequence
PKA	R-R-X-S/T-φ
CKI	pS-X-X-S/T



Where φ means hydrophobic, pS means phosphorylated serine, X is any amino acid, and the rest are regular amino acid symbols. All the real-world restrictions that you might expect come into play; phosphorylation sites buried in the middle of a folded protein are less likely to be phosphorylated, given the energetic and entropic cost of accessing it. Note that alpha helices, for example, don't fit into the active site of kinases. Do they get phosphorylated? Yes, though infrequently. Thermal fluctuations are enough to free up otherwise trapped targets, even if they don't happen frequently. All our statements at this scale are probabilistic.

Tyrosine kinases tend to use separate specificity-inducing domains, such as SH2 and SH3, that recognize sequences and bring the kinase to the area to act on local tyrosines (they are linked to the kinases by flexible linkers). S/T kinases, on the other hand, have protein domains on the same blob that provide recognition. Some of them use partners; the most prominent examples are cyclins. That's where cyclin-dependent kinases, CDKs, get their name. CDKs can only work when bound to cyclins.

Note that some of the sequences recognized by kinases include already-phosphorylated amino acids. They have to be activated by other kinases, thus creating a local AND gate that can regulate the order of events.

### Protein-DNA interaction

Specificity is determined by H-bonds in the major groove. H-bonds are crucial because they are directional. Hydrophobic interactions are mostly direction-independent, but H-bonds are really

energetically sensitive to alignment. Proteins, then, can "count" the number of H-bonds they encounter in the major groove and determine where to bind.

There are also sequence-specific interactions with the minor groove. Arginine can detect A/T-rich sequences in the minor groove. Lysine, on the other hand, is rarely found interacting with the minor groove. The reason is a difference in the favorability of the exchange reaction.

There are also some protein families that can recognize very different sequences. A great example are TAL effectors. Their secret is having lots of repeating sequences that each recognize a different element in their target (say, nucleotides in their DNA target).

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>.