# November 9, 2012

The work in this folder is a continuation of the work in the "zdiffs" folder. The goal is still to evaluate the accuracy of homology models made from sequence alignments of HHOMP clusters. However, there will be some changes:

* No longer will I be assigning to each residue in the prediction target (the protein whose structure we are attempting to find via homology modeling) a number equal to the z component of the distance between the predicted and actual positions of its Cα. Instead, each target residue will get a True or False value, representing whether it was paired to the same position in the template as TopMatch paired it to with a structural alignment.
* The sequences of the target and template will be the full sequences taken from the PDB entry, not the partial sequences extracted from the PDB file. In the zdiffs folder, anything not resolved structurally was also missing from the sequence, which is silly. Ideally I would like to use the sequence of the natural protein, in case there are sections removed from the gene for which they solved the structure, but I'm probably not going to do that.
* Alignments will be produced with both ClustalW using the Gonnet series, and ClustalΩ, so that I can compare them.

There will also be a few changes of form with no change in content, small fixes:

* Alignment file's names will have file extensions
* Alignments will be in FASTA format rather than Clustal format, since this is what Daniel's backend takes.
  + Side note - this makes me happy that I use BioPython. If I wrote the backend, I would just change a single word (AlignIO.read(filename, "~~fasta~~ clustal")), or add a line of code that looks at the file extension and chooses an appropriate format! This would be *inconsequential*. It really makes me wonder what things I spend time on that I wouldn't have to pay attention to with the right piece of software...