

Laboratory Worksheet, Monday, Oct. 24.

I. Protein Family Profiles I. Training Exercise. This exercise will be about constructing (“training”) a protein family profile HMM from real data. In this exercise, you will be given a sequence accession number (NP_000671: alpha1 adrenergic receptor). You will pass through some relatively simple steps: BLAST your sequence. Choose a handful of the best hits, but don’t choose overlapping sequences (choose distinct species, if possible). Then submit these protein sequences to CLUSTAL for MSA (multiple sequence alignment). You may do this from the command line using the local installation. Then use the MSA of your “seed” sequences, running this through *hmmbuild*, the profile HMM construction program in the HMMer suite. Having done this, you can compare to what Pfam has made of your sequence and its relatives.

II. Protein Family Profiles II. Aligning to the Pfam Database. This time you will be given another accession number (NP_002831: protein tyrosine phosphatase receptor). Pull out this protein and use it as input for a BLAST-like search against the Pfam database. This protein will be a complex, multi-module protein. The response of Pfam should show you several local matches. Interpret these compared to your original protein. What consequence might this have for annotating sequence from a genome wide sequencing project?

III. PHX Data. You should find PHX project data in the 548 Resources page, or on Ctools (depending on how far Dr De Wet and I have gotten with webDAV!). Please begin checking this for completeness. This is hand gathered data: you may want to improve it. Compare it to the Karlin-Mrceck paper if possible to see how complete any of these classes is compared to what was used in KM. The final point is to choose two organisms we can use in our project, and so you should be evaluating whether you want to choose organisms already begun or start over in data collection yourselves. We should ideally form two groups of two each to do this project. You probably won’t finish this this afternoon.

Comment: As a write up for this week, please copy from the screen your Clustal alignment from problem 1, and the list of families from problem 2 which related to small, structural elements or modules of your query protein.