

## Laboratory Worksheet, Tuesday, March 4.

### I. The Actin-HSP link.

Hopefully you managed to get your seed joint Actin/HSP seed alignment done. Next, let us submit this alignment to an HMM server to make a profile HMM out of it. The lab also claims to have HMMer (pronounced “hammer”), a suite of programs by Sean Eddy’s lab, for constructing such HMM’s, but I haven’t tried using that installation yet. You might want to use that for more control. If not, there are at least two sites where you can access a server to produce HMMer profiles:

(1) FIRC-IFOM in Milan: Italian Foundation for Cancer Research’s Institute of Molecular Oncology. (We can’t all use this one simultaneously!)

<http://bio.ifom-firc.it/HMMSEARCH/>

WARNING: the download of complete results will only be to a PC or Mac. You will probably have to work around this.

The Institut Pasteur in Paris also has a server:

<http://bioweb.pasteur.fr/seqanal/motif/hmmer-uk.html>

This page is a menu of HMMer programs. You can submit a m.s.a. and receive a profile HMM.

Haussler’s group at UC Santa Cruz has a server for SAM (the “Sequence Alignment and Modeling” suite parallel to HMMer), with latest version SAM-T02 running now. There are some useful bells and whistles available there, too, such as alignment logos. Access it at:

<http://www.soe.ucsc.edu/research/compbio/HMM-apps/HMM-applications.html>

Actually, if you want to submit a seed alignment, you will have to submit to the older, and less accurate, unfortunately, T99 version of SAM:

<http://www.soe.ucsc.edu/research/compbio/HMM-apps/T99-query.html>

Be sure to submit your data in the proper form. SAM-T99 uses a2m alignment format. It is explained carefully on a link from the query page. Save time and space by disabling parts of a query dealing with secondary structure, for example; do get the HMM as part of the output.

If all this works, where we are heading is to make a fused actin/HSP profile model and try to assess it for significance. If you succeed in getting a profile HMM today, you can stat using it to generate a simultaneous alignment of the entire Pfam seed group for actin and HSP. If you have time today, try beginning work on this.

**II. Codon Usage Bias Project.** (a) If you haven’t gotten done the code asked for last time, you can check the 548 resource page for modules that might be useful for this calculation. One is Bfn.pl, which calculates the bias function from Karlin-Mrázek for a

pair of distributions over codons for each amino acid family. Also, if you were unable to download data from the Karlin-Mrázek references, or couldn't program some component of the task, we will discuss this today.

(b) If you have already succeeded to do everything up to this point, then try automating the download from the sources of the sequence data. If you do not know how to write such a script, let us try to set a time for another Perl tutorial.

**III. Some extra tools.** There are some interesting alignment “editing” tools available on the web. The two I have in mind at the moment are: motif logos which produce visual summaries of the entropic profile of a piece of multiple sequence alignment. One is available at <http://www-lmmb.ncifcrf.gov/~toms/sequencelogo.html>. Actually, examples and documentation are available there, and the logo maker is available via a link.

Here are a couple of sites which explain how to have blocks of your m.s.a., as in part 1, analyzed and graphically presented in logo format. We should discuss the theory behind these sometime.

[http://bioinformatics.weizmann.ac.il/blocks/help/about\\_logos.html](http://bioinformatics.weizmann.ac.il/blocks/help/about_logos.html).

<http://www.cbs.dtu.dk/~gorodkin/appl/plogo.html>.