

## Laboratory Worksheet, Tuesday, April 2.

**0. Basic Lab Procedure** Start a new directory entitled 548\_0402. Please keep in it the results of any computations which you do today which you do not keep in topic-titled files like “actin-HSP” or “Codon usage bias”, for example.

Recall that you have access to accounts on the math lab machines in East Hall. In particular, the math labs all have Matlab mounted, and have the neural nets toolbox with a reasonably friendly user interface. If I can get the software required to work on a simple but illustrative neural net problem by next week, we will have our lab there next week. You will be notified of this by email. **Please check to see whether you still have the initial math labs account passwords distributed the first day of 548.** If you cannot find yours, or never received one, and are currently registered for 548, write an e-mail to math-req@umich.edu explaining your situation. If you are not registered, remind me today and I will arrange an account. (Remind me also of the unickname you use at the BICC.)

**I. tRNA Synthetases, part I.** This is to continue our third project, on tRNA synthetases. What you should have now is a (nearly) complete file of sequence data for all the aminoacyl-tRNA synthetases (or acyl-RS's) for *E. coli*. Recall that these are the enzymes which charge the various tRNA's with the appropriate amino acid that “matches” the tRNA's codon. Thus they represent the specificity of the genetic code, and are central to the establishment of a world of proteins synthesized by nucleic acids. They are very conserved within their types across species, but the relationships among them are quite complex, it seems. As mentioned last time, there is an important breakdown of acyl-RS's according to “class”: I and II. These are markedly distinct in their homologies, and that should be showing up in our alignments. There are recent comments on this in Carter-Duax, letter to *Mol. Cell*, **10** (2002), 705-708. The classification is discussed in Eriani, et al., *Nature*, **347** (1990), 249-255, and Cusack, et al., *Nature*, **347** (1990), 203-206. Their homologies to one another are at times tenuous (especially among the class II acyl-RS's). Structural similarities seem necessary to see the relationships among them. To get a brief outline of this classification online, with useful links to Medline abstracts, go to the cover pages for the Pfam listings for the families of class I tRNA synthetases (PF00133) and class II tRNA synthetases (PF00152).

From last time, you should have:

- 1.) A small data file of acyl-RS sequence from GenBank, or some other source, for *E. coli*. Recall that a set of 16 such are available at the 548 resources directory.
- 2.) You should have multiple sequence alignments for the data in 1.) above.
- 3.) Did any of you get to make a phylogeny of these sequences last week? Can you pick outliers, or estimate relative ages of the RS's?

To start with today:

- 1.) Partition your acyl-RS file into two files, with the same sequence data, one for class I acyl-RS's and the other for the class II acyl-RS's. For convenience, these classes are reproduced at the end of the worksheet below.
- 2.) Perform the same analyses for the two split files as for the single file.

It would be best to use SAM at

<http://www.cse.ucsc.edu/research/compbio/HMM-apps/T-99-query.html>. The alignment methodology here adds structural information from databases as it computes the alignment, and would be better for our problem. This takes more computational time, however. One could get a seed alignment from a ClustalW program, and tune it up using <http://www.cse.ucsc.edu/research/compbio/HMM-apps/T99-tuneup.html> and HMM training. I think, however, that our alignments will be small enough that we can submit the whole file directly to the “tuneup” utility for alignment. The input for the “query” function should be an alignment.

**II. tRNA Synthetases, part II.** In this part of the exercise, let us now go back to the databases and assemble another small data set. This time I want you to collect sequence data for tryptophanyl-RS *across* several species. (If you are efficient, do the same for tyrosyl-RS and phenyl-RS also, cf. below.) Some of them should be distant species, probably including some prokaryotes and archaea. Again, make a multiple sequence alignment. For the reasons outlined above, it might be good to use the SAM modeler so that structural information is taken into account. The question I want you to test is: *does the consensus sequence you get for your alignment contain any conserved tryptophan residues?*

The idea is that according to certain laboratory simulation experiments, the Miller-Urey experiments of the 50's and their descendents, there seems to be a partition of amino acid residues into those which are primitive and those which are more complex and more recent. The first group can be generated “spontaneously” from inorganic settings. (There have been claims of “abiotic synthesis” of each of the twenty amino acid residues used in proteins, however.) Last week we were looking for a cluster apart for the “older” amino acids. This probably is not visible. There are, however, researchers who currently are searching to separate older from younger amino acids by consensus sequences of the corresponding acyl-RS's. In fact, they are finding consensus sequences for later amino acids' acyl-RS's which do not contain the later amino acid residues, the idea being that perhaps there are traces of a historical succession, whereby older amino acids might have formed acyl-RS's before the incorporation of the later amino acids into the repertory of proteins. At any rate, the last I heard, they were trying to synthesize these consensus sequences to see whether they would be functional as acyl-RS's (!).

If you finish this tryptophan exercise quickly, you may try the same thing with tyrosyl-RS and phenyl-RS. Note that these two amino acids are relatively complex, but one is of class I (pheRS) and one of class II (tyrRS).

## Class I acyl-RS's

Arginyl-RS (AgnRS)  
Cysteinyl-RS (CysRS)  
Glutamyl-RS (GluRS)  
Glutaminyl-RS (GlnRS)  
Isoleucyl-RS (IleRS)  
Leucyl-RS (LeuRS)  
Methionyl-RS (MetRS)  
Tyrosyl-RS (TyrRS)  
Tryptophanyl-RS (TrpRS)  
Valyl-RS (ValRS)

## Class II acyl-RS's

Alanyl-RS (AlaRS)  
Asparagyl-RS (ApnRS)  
Aspartyl-RS (AspRS)  
Glycyl-RS (RS)  
Histidyl-RS (HisRS)  
Lysyl-RS (LysRS)  
Phenyl-RS (PheRS)  
Prolyl-RS (ProRS)  
Seryl-RS (SerRS)  
Threonyl-RS (RS)